Our Ref.: SP27447OP **D8**

Declaration of Michael A. Frohman, MD, PhD

I, Prof. Michael A Frohman, declare and state as follows:

I. Background

1. I am a SUNY Distinguished Professor and the Chair of the Department of Pharmacological Sciences at the Stony Brook University Renaissance School of Medicine in Stony Brook, New York. I held the position of Director of the Molecular Cloning Facility Core at Stony Brook University from 2002 to 2017. Since 2003, I have also been the Director of the Medical Scientist Training Program. A copy of my *curriculum vitae* including a publication list accompanies this declaration.

2. As a professor and principal investigator, my major interest lies in unravelling translational roles for members of the mammalian family of lipid-signaling Phospholipase D genes, as can be seen from the more than 200 peer reviewed original papers and reviews I have authored. My areas of expertise also include biochemistry, immunology, developmental and molecular biology, with 38 years of experience in molecular cloning. In particular, I developed the well-known RACE protocol (which was published in the *Proceedings of the National Academy of Sciences* in 1988) during my Postdoctoral studies on homeobox-containing genes at the University of California at San Francisco, which has since then greatly facilitated cDNA cloning and, nowadays, cataloging of endogenous RNAi. This publication, which was designated a "citation classic", and ones related to it have been cited more than 8000 times.

3. Consequently, I believe that I am qualified to provide an opinion as to what a person of ordinary skill in the art would have known and concluded as of the claimed priority date of EP 3 530 670 B1 (October 18, 2016).

II. Instructions

4. I am familiar with European Patent EP 3 530 670 B1 (hereinafter referred to as "EP'670") for which claim 1 relates to a method for producing an endotoxin detecting agent comprising a step of artificially expressing a protein using a cDNA [...], wherein the cDNA encodes [a limulus Factor C sequence]...

5. I have been asked by the Opponent's representative to review the experimental section of EP'670 and to provide an opinion on the following questions from the perspective of a skilled person charged with the determination of the full-length *Limulus* Factor C sequence **before October 18, 2016**:

   (a) How would the skilled person wishing to obtain the coding sequence of the *Limulus* Factor C protein have gone about obtaining this information?

   (b) How would the skilled person have designed the primers needed to carry out RACE?

(c) Would the outcome of the cDNA sequence determination depend on the specific RACE protocol used?

(d) Are there any technical difficulties one might encounter when performing RACE, which might have made determining the *Limulus* sequence challenging?

(e) Is there any indication in the patent (EP'670) that a technical difficulty had been encountered when determining the *Limulus* sequence in the examples?

(f) Is there any reason to believe that this protein would have been difficult to express after the cDNA sequence had been obtained?

6. I have further been asked to include the following information in my analysis, which was publicly available before October 18, 2016:

(i) Genomic data for *Limulus polyphemus* that was publicly available in the NCBI database as of January 3, 2014 (see the NCBI entry in **D18**);

(ii) An mRNA sequence encoding part of the *Limulus polyphemus* Factor C (hereinafter referred to as "*Limulus* Factor C") that had been predicted by an automated computational analysis in the NCBI database under the accession number XM_013937646.1 (see **D20**);

(iii) Partial amino acid sequences of the *Limulus* Factor C protein that were available in the publication by Tokunaga (**D22**, page 157, Figure 7S); and

(iv) A cloning protocol for obtaining the cDNA sequence of the *Carcinoscorpius rotundicauda* Factor C (hereinafter referred to as "*C. rotundicauda* Factor C") that was described in US 5,7168,834 (**D23**) and the publication by Ding, 1995 (**D24**).

III. How would the skilled person wishing to obtain the coding sequence of the *Limulus* Factor C protein have gone about obtaining this information before October 18, 2016?

III.1 *The skilled person would have known to use RACE to obtain the necessary sequence information*

7. In 2016, the skilled person would have most likely used Rapid Amplification of cDNA Ends (RACE) to obtain the sequence of the cDNA encoding Factor C from *Limulus polyphemus*.

8. I developed RACE in 1988 as a quick approach to obtain full-length cDNA sequences in cases where a partial stretch of the mRNA sequence was known. The original paper disclosing this method is provided as "**D12**". In contrast to the more traditional approach of screening cDNA libraries and selecting positive clones, which is very time intensive, RACE can be used to obtain a full-length cDNA sequence in a matter of days. Given the ease of using RACE, it quickly gained attention in the field - in 1989 alone, it was cited 9 times in renowned scientific journals such as Science, Nature and Cell. So, the protocol was well-recognized within a very short period of time.

9. By 2016, **D12** had been cited over 5,000 times according to Google Scholar and kits had been available for more than 20 years. At this point, RACE was considered the "go-to" method for this purpose and would have been the obvious choice for someone wanting to determine the cDNA sequence of the *Limulus* Factor C.

10. It would have been entirely straightforward for the skilled person to design RACE primers based on the available *Limulus* sequence information listed in items (i)-(iii) above and then use any of the available RACE-PCR kits to obtain the full-length cDNA sequence.

III.2 *The skilled person could have followed the teaching described in example 1 of **D23** (US 5,716,834), or equally **D24** (Ding, 1995)*

11. Each of these disclosures describes obtaining the cDNA sequence encoding Factor C derived from *C. rotundicauda*.

12. In **D23** and **D24**, mRNA extracted from *C. rotundicauda* horseshoe crab blood is used to generate cDNA. Adapter sequences were added to the cDNAs to generate restriction sites at the ends and the cDNAs cloned into λGT22. The resulting recombinant phage DNAs were then transduced into *E. coli* to produce a library. A cDNA probe encoding *Tachypleus tridentatus* Factor C (hereinafter referred to as "*T. tridentatus* Factor C") was then used to screen this library, whereby a selection of presumptive clones was identified, isolated and further characterized. In particular, the sizes of their inserts were determined and a subset was sub-cloned, transformed into *E. coli,* and characterized by restriction mapping and DNA sequencing.

13. The approach described in **D23** and **D24** was successful because the cDNA sequences encoding the Factor C coding regions of *T. tridentatus* and *C. rotundicauda* were similar enough for hybridization approaches to be usable. As may be taken from the sequence alignments in Annex I, the cDNA encoding the Factor C derived from *T. tridentatus* is 97.1% identical to the cDNA encoding the Factor C derived from *C. rotundicauda*. Given the high level of identity between these sequences, the *T. tridentatus* sequence was adequately able to serve as a hybridization probe.

14. The protocol disclosed in **D23/D24** is very straightforward as described above. There's every reason to believe that this approach would have been successful to obtain the *Limulus* Factor C cDNA. Given the success of obtaining the cDNA sequence of the other, closely related species, there would be no reason to expect any technical difficulties for repeating this protocol for *Limulus*.

15. If we look at the cDNA sequence encoding the Factor C derived from *Limulus,* we see that it is 90.1% identical to that of *T. tridentatus* and 89.3% identical to the cDNA sequence of *C. rotundicauda*. Given the high level of identity between these sequences, either the *T. tridentatus* or *C. rotundicauda* cDNA sequence could have been used as a hybridization probe to screen the *Limulus* cDNA library and would have been expected to work.

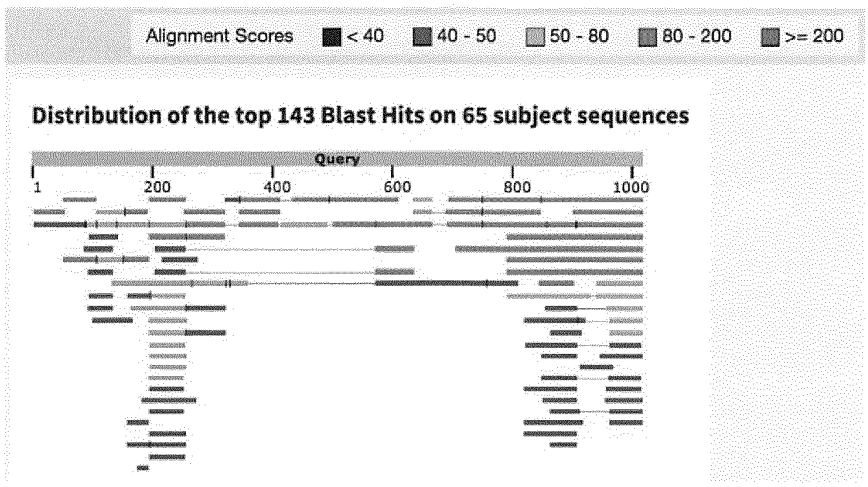
16. So, in my opinion, the experiments described in **D23/D24** could have been repeated to obtain the *Limulus* Factor C sequence in a straightforward manner. Although as explained above, I believe the skilled person would have favored the less time-intensive PCR-based RACE approach in 2016.

IV. How would the skilled person have designed the primers needed to carry out RACE?

17. I personally would have started with the genomic information in the NCBI database (item (i) above). However, any of the sources of information listed in (i)-(iii) above could have been used to design RACE primers. As will become apparent from the discussion below, the amount of sequence data that was publicly available at the time provided more than enough information to obtain the single, short gene-specific primer minimally needed to carry out RACE.

*IV.1 Starting from the genomic data (**D18**)*

18. By 2016, the *Limulus polyphemus* genome was publicly available in the NCBI database. Using tBlastn with the published *T. tridentatus* Factor C protein sequence[1] against the collection of *Limulus* shotgun genome fragments returns the information shown in Appendix A[2]. A graphical summary of the results is shown in the figure below:



19. In brief, almost the entire *Limulus* coding region is identified, although the sequence at the very 5' end is not certain. This provides internal, known, coding sequence regions for the *Limulus* protein

[1] The *T. tridentatus* Factor C sequence was published in, e.g., **D1** (SEQ ID NO:2); **D23** (See Figure 11); **D24** (see Figure 5) and was further publicly available in the NCBI database as of Feb. 6, 2002 (see Appendix D)

[2] tBlastn is a search tool provided by the NCBI (National Center for Biotechnology Information) site, which allows a user to search translated nucleotide databases using a protein sequence as a query. Here, the *T. tridentatus* protein sequence was used as query to search for similar sequences within the database containing the *Limulus* genomic information. The search returned "hits" from the *Limulus* database and ranked each hit (i.e., *Limulus* genomic sequence) according to its overall goodness of fit to the *T. tridentatus* protein sequence. These are listed in Appendix A and shown graphically above.

and the starting place for using RACE-PCR. It should be noted that the alignment scores are variable in quality, reflecting differing levels of error in the shotgun genomic sequences.

20. The comparisons at the amino acid level for the first 3 exons are shown below:

```
Exon 1
T. tridentatus     6  FLVSGLVLGILAQQMRPVQSRGVDLGLCDETRFECKCGDPGYVFNVPMKQ  55
Limulus genome (4)    FTFPGLVLGLLAQQMHPVQSRGVDLGLCDDTRFECKCGDPGYVFNVPAKQ
Limulus protein       FLVSGLVLGLLAQQMHPVQSRGVDLGLCDDTRFECKCGDPGYVFNVPAKQ

Exon 2
T. tridentatus    56  CTYFYRWRPYCKPCDDLEAKDICPKYKRCQECKAGLDSCVTCPPNKYGTWCSG 108
Limulus genome (1)    CTYFYRWRPYCKPCDKLEAKDVCPKYKRCQECRAGLDSCVICPPNRYGTWCSG
Limulus genome (2)    CTYFYRWRPYCKPCDKLEAKDVCPKYKRCQECRAGLDSCVSCPPNKYGTWCSG
Limulus protein       CTYFYRWRPYCKPCDKLEAKDVCPKYKRCQECRAGLDSCVSCPPNKYGTWCSG

Exon 3
T. tridentatus     109  ECQCKNGGICDQRTGACTCRDRYEGAHCEILK  139
Limulus genome (1)      ECQCKNGGICDQRTGACTCRDRYEGVHCEILQ
Limulus genome (3)      ECQCKNGGICDQRTGACTCRDRYEGVHCEILQ
Limulus protein         ECQCKNGGICDQRTGACTCRDRYEGVHCEILQ

Limulus genome (1) NW_013681153.1
Limulus genome (2) NW_013850675.1
Limulus genome (3) NW_013820061.1
Limulus genome (4) NW_013690523.1
```

Legend: The *T. tridentatus* protein sequence used as the query is shown in comparison to the highest-quality translated regions of the *Limulus* genome fragments listed. Red text indicates identity. Blue, underlined text indicates differences between the two genomic fragments in Exon 2. The full *Limulus* protein, from EP3530670, does not appear to have been publicly available before October 18, 2016 but is shown for comparison. In most but not all cases, the differences between the translated *Limulus* genomic regions and the *T. tridentatus* protein sequence were similarly different in the *Limulus* protein sequence.

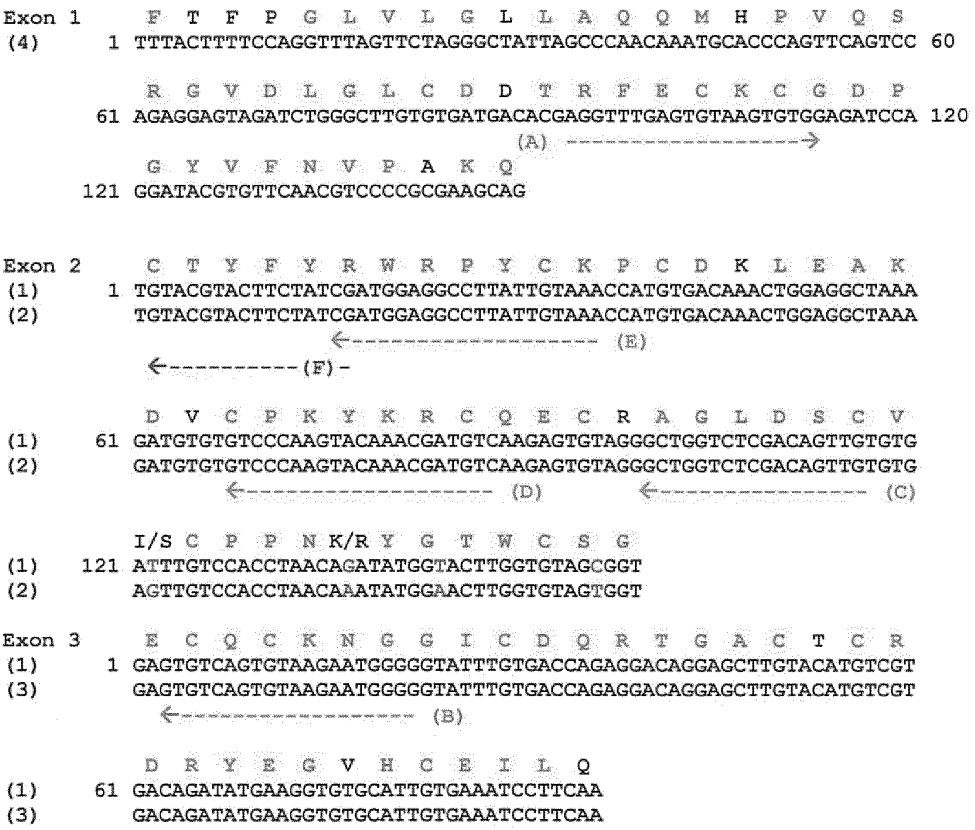
21. As may be taken from the above, most of the protein-encoding sequence was available – but the sequence corresponding to the first ~9 amino acids of the *Limulus* protein was not obtained or certain from the tblastn search and would need to be generated using 5' RACE[3].

22. Primers for 5' RACE could have been designed based on the DNA sequences of the first 3 exons, as shown below[4]:

[3] In Appendix B, the exons identified in the tBlastn search have been overlaid with the *Limulus* protein sequence from EP3530670 (SEQ ID NO 2). As may be taken from this analysis, the tBlastn search provides the majority of the *Limulus* sequence. When the exon information is combined with the predicted Factor C sequence from the database (**D20**), **95%** of the *Limulus* protein-encoding sequence is obtained with confidence. Thus, 5' RACE would have only been necessary to determine the remaining 5% of the protein-encoding sequence.

[4] The proposed primer design scheme utilizes regions of amino acid identity to the *T. tridentatus* protein sequence (indicated by red amino acids) to maximize chances for success.

Seikagaku Corporation Exhibit 2004 - Page 5 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023



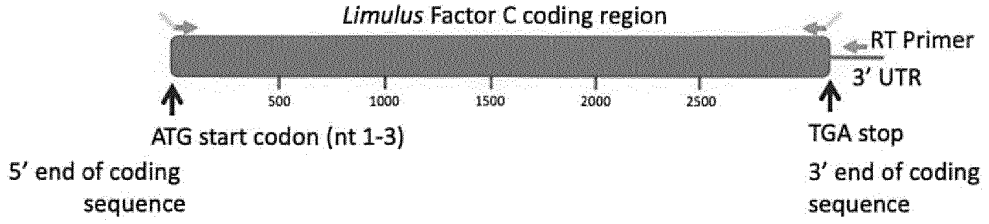
23. Primers would be generated according the protocol recommended by the 5' RACE kit used. In general, reverse transcription of RNA prepared from blood cells of *Limulus* would be carried out using Primer C. The resulting cDNA pool would then be tailed/anchored as described in the kit and then nested RACE carried out using Primer D and the initial tailed anchor RACE primer. An aliquot of the amplification products would then be reamplified using the nested Primer E and the non-tailed RACE primer.

24. For an exemplary 5' RACE kit, please see **D16** (the instruction manual accompanying the Invitrogen 5' RACE Kit dated 2004). In this kit, the initial tailed anchor RACE primer and the non-tailed RACE primer described above would correspond to the "abridged anchor primer" and the "AUAP" (or "UAP") in Figure 1, respectively.

As can be observed in the sequences above, there is a small frequency of polymorphism in the genomic fragments, in particular at wobble codons. Given that the RNA prepared would be from a different, wild-type horseshoe crab, the possibility for a single nucleotide variation exists at any of many positions and *could* prevent the RACE primers from being extended. This is unlikely, but possible. If problems were encountered, one could first use primers A and B or similar primers to amplify the sequence between them from reverse-transcribed RNA from blood cells and then sequence the amplified material; this would provide the exact sequence in the individual horseshoe crab being used to confirm or change the primers used for 5' RACE.

25. An aliquot of the resulting amplification product would be electrophoresed on a 1% agarose gel along with a DNA ladder and the DNA species visualized using ethidium bromide. Given the position of Primer E in the *T. tridentatus* Factor C sequence, one would anticipate a reaction product of 201 bp, plus the unknown length of the 5' UTR and the length of the 5' RACE primer appended. Presuming that an appropriate-sized band of DNA is observed, another aliquot of the amplification product would be sequenced using **Primer F** to obtain a validated sequence at the 5' end of the coding region.

26. **Generation of a full-length coding sequence:** Once the 5' end of the coding sequence was known, primers would be generated to amplify the entire coding sequence along with appended restriction sites for sub-cloning into the vector of choice as shown in the figure below. *Limulus* blood cell RNA would be reverse transcribed with an RT Primer chosen just 3' of the stop codon and then PCR carried out on the resulting Factor C-enriched cDNA pool using a high-fidelity heat-stable polymerase. After confirming with gel electrophoresis that a DNA amplification product of the expected size had been generated, the amplification products would be cleaned with a DNA spin column to remove primers, nucleotides, and heat-stable polymerase, digested with the restriction enzymes needed for sub-cloning, and gel electrophoresed. The desired DNA band would be excised from the gel and processed to recover the DNA for ligating into a prepared expression vector. The ligated plasmids would be transfected into *E. coli* and individual colonies grown to prepare plasmid DNA for sequencing. A plasmid encoding a sequence with fidelity to the sequences earlier defined would be chosen for protein production.



IV.2 *Starting from the known, predicted Limulus sequence (**D20**)*

27. A 3' partial sequence for *Limulus* Factor C had been identified by an automated computational analysis of a cDNA deposited in NCBI on September 9, 2015 in GenBank (XM_013937646.1).

28. This would have been another good starting point for obtaining RACE primers. One could have mapped this sequence onto the known *T. tridentatus* or *C. rotundicauda* sequence to confirm the 3' end of the coding sequence; they would have then only had to perform 5' RACE to obtain the missing end. The open reading frame for this sequence ends with a stop codon at the same place in which a stop codon is present in the *T. tridentatus* sequence, validating its identification as the carboxy-terminus of the protein.

29. Once 5' RACE had been carried out to complete the sequence acquisition, the skilled person would then amplify the full-length open reading frame and clone this product into a vector which would then be introduced into *E. coli*.

IV.3 *Starting from the disclosure of **D22** (Tokunaga)*

30. **D22** provides peptide sequences in 3 different domains of the protein (see figure 7S on page 157). One could, in principle, use the *Limulus* amino acid sequences shown in Figure 7S to design degenerate primers that could be used to clone a partial cDNA fragment from H to B using *Limulus* cDNA. Sequencing this product would provide the sequence information needed for RACE-PCR.

31. Alternatively, these peptides could be used for a tblastn search against the *Limulus* genome. As shown in Appendix C, this approach is successful for the H chain peptide, and results in the identification of 138 nt of coding sequence in exon 1 (as defined elsewhere). This sequence would more than suffice to act as a hybridization probe in a screening experiment; or to generate primers for 5' and 3' RACE.

32. 3' RACE and 5' RACE could then be performed with any of the many available kits to determine the ends of the cDNA sequence.

33. Although this approach should also be possible, I believe the skilled person would have favored the genomic data or the predicted Factor C sequence as the starting point for RACE in 2016 since they are richer sources of information and would only require the use of 5' RACE. In particular, and as noted above, most of the *Limulus* Factor C coding sequence can be derived directly from the genomic data with a high level of confidence, so this is where I would start for RACE since it would involve the least amount of effort.

V. Would the outcome of the cDNA sequence determination depend on the specific RACE protocol used?

34. No.

35. Since RACE was initially developed, there have been various modifications to the original protocol (PMID: **7573945**). However, these modifications merely optimize the method for efficiency. For example, different RACE strategies have been developed to minimize the number of steps in the method. So, one protocol might be easier to perform than others and may be a bit faster, but ultimately, they are all able to provide the necessary information for expression cloning.

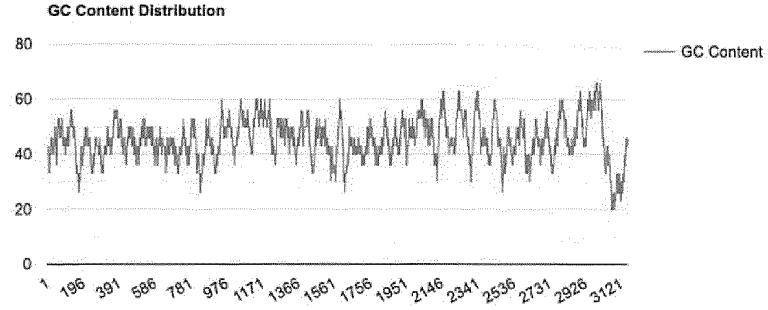
36. In particular, mRNAs from animals have 5' UTRs that are typically on the order of 100 nt or more. For the purpose of expression cloning, one just needs the ORF sequence; knowledge of the exact sequence at the very 5' end of the mRNA is not needed- any of the known RACE methods could be used to determine the ATG translation start site.

37. There were certainly kits available in 2016 to perform 5' RACE, and any one of these could have been used to determine the full-length *Limulus* cDNA sequence. It merely would have been a matter of convenience/preference as to which kit the skilled person used.

VI. Are there any technical difficulties one might encounter when performing RACE, which might have made determining the *Limulus* sequence challenging?

38. One might face some difficulty in determining the cDNA sequence if it had regions in which the GC-content exceeded 80%, because of secondary structures. However, I would not have expected to encounter this issue based on the known sequences of *T. tridentatus* and *C. rotundicauda*.

39. Shown below is a GC content analysis of a *C. rotundicauda* Factor C cDNA sequence. The overall GC content is 45% and the GC content does not exceed 60% until past nt 2000. 5' RACE PCR would have been performed on the first 300 nt. So, this would not have been an issue.



Summary: Full Length(3180bp) I A(27% 866) I T(28% 856) I G(25% 799) I C(20% 659)

VII. Is there any indication in the patent (EP'670) that a technical difficulty had been encountered when determining the *Limulus* sequence in the examples?

40. No.

41. Aside from some gratuitous "Surprisingly..."s, there's no sign that any of it was hard to do or required modifications of the RACE protocol that would have posed challenges to others attempting the procedure.

42. The example describes standard 3' and 5' RACE protocols, followed by standard cloning to obtain the full-length cDNA sequence.

43. I do not see that the authors describe encountering any substantive issues while performing this experiment; I do not see any steps being described which may be considered unusual for RACE.

VIII. Is there any reason to believe that this protein would have been difficult to express after the cDNA sequence had been obtained?

44. No.

45. Again, the cDNA sequence of the *Limulus* Factor C is very similar to the other known species of horseshoe crabs, which had been successfully expressed in Sf9, CHO DG44 and HEK 293 cells. This has been shown in e.g., **D1** ('241).

46. Given that these cells were able to incorporate vectors containing cDNA for *T. tridentatus* and/or *C. rotundicauda* and express the Factor C-encoding mRNAs and Factor C proteins, there's no reason to think that they would not be equally capable of handling the *Limulus* sequence.

47. If I were to follow the example of **D1** using the *Limulus* sequence, I would start with the cells exemplified in this disclosure and would not anticipate experiencing any difficulties.

48. I declare that the foregoing statements are true and correct, to the best of my expert and personal knowledge.

Executed on November 14th, 2024 in Stony Brook, NY, USA.

Prof. Michael A. Frohman (MD, PhD)

Enclosures
Appendices A-D
Curriculum Vitae

MICHAEL AARON FROHMAN

Born May 18, 1958
Ann Arbor, Michigan

Work Address: Center for Developmental Genetics,
438 CMM
Stony Brook University
Stony Brook, NY 11794-5140

Phone #: 632-1476; e-mail: Michael.frohman@stonybrook.edu

A. Personal Statement

My laboratory cloned the mammalian family of lipid-signaling Phospholipase D genes and has worked on them for the past 29 years while being continuously funded from the NIH and publishing more than 100 peer-reviewed articles and reviews on PLD and other lipid signaling topics with additional ones in submission. PLD superfamily members are involved in many physiological and pathophysiological settings including immune mitochondrial biology and cardiovascular disease. Among other approaches, we have been generating and publishing findings with mice lacking each of the PLD isoforms and have uncovered a number of fascinating stories with human health relevance. Recent work has included exploring the potential of using a Phospholipase D small molecule inhibitor as a therapeutic in stroke and cancer progression; additional manuscripts on the cancer-focused topic are currently in revision.

I have been the Chair of Pharmacology for 17 years, hiring and mentoring 17 faculty during this period. I have been Director of our Medical Scientist Training Program (MSTP, MD-PhD) for 21 years, renewing the T32 training grant 4 times and graduating 144 trainees. I have trained >50 students and fellows in my own lab, ranging from high school students and undergraduates to graduate and MSTP students, postdoctoral fellows, and research faculty. Approximately 20% of my trainees have been/are URM. I have been an external reviewer of Graduate / MD-PhD training programs since 2006 and have participated in national conferences, study section reviews of MSTP T32 training grants and F-series NRSAs.

As a Chair, NIH T32 training grant Director, and lab PI, I am deeply committed to promoting an inclusive, safe, and supportive research environment. Throughout my career, I have been committed to maintaining a record of and providing training in rigorous and unbiased experimental design, methodology, analysis, interpretation, and reporting of results. I am also committed to supporting my trainees participating in activities required to identify and transition into careers in the biomedical research workforce that are consistent with their skills, interests, and values. Almost all of my former trainees are engaged in science-connected careers. Finally, I am committed to helping my trainees to complete their Ph.D. degrees in a timely fashion with the skills, credentials, and experiences needed to transition into careers in the biomedical research workforce.

EDUCATION

1975-1978	B.S. (Chem.) with Distinction, High Honors in Chemistry and High Honors in Cellular & Molecular Biology University of Michigan, Ann Arbor, MI
1979-1985	M.D., PhD., Immunology (MSTP). University of Pennsylvania Medical School, Philadelphia, Pennsylvania

UNDERGRADUATE RESEARCH

Summers, 1976, 1977	University of Chicago, (laboratory of Dr. John Goldman). Research area: novel rat complement fixing factors, development of radioimmunoassay for fifth component of complement

Academic year, 1977-78	University of Michigan, (laboratory of Dr. William Kelley). Research area: Adenosine deaminase, adenosine deaminase binding protein.
January - May 1979	Research technician, Michael Reese Hospital, University of Chicago, (laboratory of Dr. Michael Berilowitz). Research area: Somatostatin biosynthesis.

DOCTORAL DISSERTATION

1979-1985	University of Pennsylvania, Department of Pathology, (laboratory of Dr. Carol Cowing) "Immunological aspects of Ia on non-classical antigen-presenting cells."

POST-DOCTORAL RESEARCH AND APPOINTMENTS

1986 - 1992	Postdoctoral fellow in the laboratory of Dr. Gail R. Martin, UCSF Research interests: Pattern formation and gastrulation in vertebrates, molecular biology of early mammalian development, homeobox-containing genes.
1992 – 1998	Assistant Professor, Stony Brook University, Dept. Pharmacology
1994 - 2003	Course Director, Medical Pharmacology
1998 - 2003	Associate Professor, Stony Brook University, Dept. Pharmacology
2002 - 2017	Director, Molecular Cloning Facility Core, Stony Brook University
2003 -	Director, Medical Scientist Training Program at Stony Brook University
2003 - 2016	Professor, Stony Brook University, Dept. Pharmacology
2007 -	Chair, Department of Pharmacology
2016 -	SUNY Distinguished Professor, Stony Brook University
2018 -	Associate Director for Education, Stony Brook Cancer Center

HONORS

1977	Michael Reese Medical Student Research Scholarship	
1978	University of Michigan Medical Student Summer Research Fellowship	
1979-1985	Scholarship: NIH Medical Scientist Training Program	
1984	Dr. Roy G. William Award	
1986	Post-doctoral Fellowships:	American Cancer Society National Institutes of Health (declined) Jane Coffin Childs Memorial Fund (declined)
1989		Leukemia Society Senior Fellowship American Cancer Society Senior (declined)
1993	Basil O'Connor Young Investigator Award, March of Dimes	
1997	Keynote speaker, Southeast Regional Lipid Conference	
2001	Dean's Research Productivity Award, School of Medicine, Stony Brook University	
2002	Dean's Research Productivity Award, School of Medicine, Stony Brook University	
2011	Keynote speaker, National conference for MD-PhD training, Gwangju, South Korea	
2015	School of Medicine Excellence in Faculty Mentoring Award	
2015	Fellow of the American Association for the Advancement of Science (AAAS)	
2016	Keynote speaker, Penn State MD-PhD program	
2016	Elected into the Association of American Physicians (AAP)	

2016	SUNY Distinguished Professor, Stony Brook University
2017-19	President, American Medical School Pharmacology Chairs association
2020	Member, National Academy of Inventors
2021	Fellow, American Society for Pharmacology and Experimental Therapeutics (ASPET)
2022 -	Ambassador for the University of Thessaloniki

ACTIVITIES

1990	Co-Chairman for (first annual) "Gastrulation Symposium" at UCSF
1994, 95	NIH Cell Biology I Study Section Special Reviewer
1996	NIH HED-2 Study Section Special Reviewer
1998	Meeting co-chair, session chair, FASEB meeting on Phospholipases, Vermont
1998 - 2003	Editorial Board, Journal of Biological Chemistry
1999 - 2002	NIH Special Emphasis Panel Study Section Ad-hoc Reviewer (ZRG1 SSS-Y 01)
2000	Symposium co-chair; "Roles of PLD in Signal Transduction" ASPET meeting, Boston
2001	Chairman (organizer), FASEB meeting on Phospholipase D, Arizona
2002 -	Director, Molecular Cloning Facility Core
2002	NIH P01 Special Emphasis Reviewer
2004- 07	Editorial Board, Journal of Endocrinology
2005-	External Advisory Board, Penn State Medical Scientist Training Program
2007-08	Guest Co-Editor, Special issue on Phospholipase D in BBA Lipids
2007-	Editorial Board, Journal of Functional Development and Embryology
2008 - 13	Editorial Advisory Board, Molecular and Cellular Pharmacology
2009	NIH MPP Study Section Ad-hoc Reviewer
2010	External reviewer, Molecular Medicine Graduate Program, Med. Coll. Of Georgia
2012 - 15	Review Editorial Board of Frontiers in Mitochondrial Research
2013 -	Internal Advisory Committee, Genetics Training Program and T32 Training Grant
2013 - 15	Councilor, *Association of Medical School Pharmacology Chairs* (*AMSPC*)
2013	NIH T32 Study Section Ad-hoc Reviewer for MSTP applications (February and June)
2014 - 22	Editorial Board, "*Handbook of Experimental Pharmacology*"
2014 -	Advisory board, DO/PhD program, New York Institute of Technology
2015	NIH NCATS Review Panel (R13)
2015 - 16	Internal SMART Program Integrated Advisory Committee (for the Dental school DDS/PhD training program)
2017 -	Member and Steering committee member of the Institute for Chemical Biology and Drug Discovery
2017 - 18	Consultant and board member, Sevengenes
2017 -	Associate Editor, Journal of Pharmacology and Experimental Therapeutics (JPET)
2018	Advisory Committee in preparation of the 2020 FASEB Scientific Research Conference on "Phospholipids: Dynamic Lipid Signaling in Health and Disease".
2019	Committee to establish authorship policies for SBU
2019 -	Committee to elect SUNY Distinguished Professors
2021, 22, 23	NIH NCI K22, R25, and T32 study section
2020 - 22	Member, ASPET Awards Committee
2022	NIH NCI Study Section review of Univ. of Arizona Cancer Center

2023 NIH NINDS T32 study section

PROFESSIONAL ASSOCIATIONS

1986 - AAAS
1994 - Society for Neuroscience
2002 - American Diabetes Association
2002 - ASPET

COMMITTEES, LOCAL

1992 - 1994 Medical School Admissions Committee
1994 - 2002 Faculty Senate
1993 Unnamed committees which established courses in Developmental Biology and Regulatory Biology (I teach in both courses)
1995 - 2003 Transgenic Animal Core Facility Committee
1997 - 1999 Basic Sciences Steering Committee (Primary task, Dental School accreditation)
1998 - 2003 Technology and Patent Review Committee
1998 - 2003 Curriculum Committee
1999 - 2003 Appointments, Promotions, and Tenure Committee
1999 Pain Management Course Development Committee
2000 - 2002 Finance Committee, Dept. Pharmacology
2000 - 2003 Chair, Transgenic Animal Core Facility Committee
2000 - Grievance committee for the graduate program in Molecular and Cellular Pharmacology
2001 - 2002 Member, Dean's search committee for Pharmacology Chair
2002 - 2003 Strategic Planning Committee for Research 2002/3, School of Medicine
2003 Chair, Genetics Qualifying Exam
2002 Advisory Board for planning research infrastructure at the School of Dental Medicine
2003 - Medical School Admissions Committee
2007 Executive committee, Dept. Pharmacology
2009 Search committee for head of Research Foundation
2009 LCME review committee on medical school admissions
2010 Search committee, Dean of School of Medicine, Stony Brook University
2011 Bain Review Group on Research Foundation structure
2011-12 Search committee, Physiology Chair, Stony Brook University
2012 Search committee, Laufer Center's Endowed Junior Chair Faculty, Stony Brook University
2012 Search committee, Deputy Director for Research, Cancer Center, Stony Brook University
2012 - Internal Advisory Board for the Stony Brook University Cancer Center
2012 - 17 Executive Core Oversight Committee
2012 - 17 Central Microscopy Imaging Center Core Advisory Committee
2012 - Institutional Research Advisory Committee
2013 - 15 Search committee, cluster hire for Bioimaging
2013 -18 Advisory Board for planning research infrastructure
2013 Academic Recharge Center Policy & Procedure Working Group

2013 - SBU Cancer Center External Relations Committee
2016 Stony Brook Discovery Prize Selection committee
2017 - Educational program, Cancer Center NCI application
2017 Bahl gift committee
2017- Member & Steering committee member, Institute for Chemical Biology and Drug Discovery
2020 - Advisory/executive committee for CTSA educational core
2022 Search committee for Vice Dean for Research
2023 Search committee for Cancer Center Director

TEACHING

Courses Directed:

- Principles of Medical Pharmacology (150 medical and dental students); 1994 – 2003
- Pharmacological Systems (a 5-week module in Grad. Program in Pharmacology, HBH633); 2000-08
- HBH560, Pharmacology Qualifier Preparation course, 2006-09

Team teaching participation:

- Principles of Medical Pharmacology (150 students): Endocrine Pharmacology; 1994 - 2007
- Principles of Medical Pharmacology (150 students): Antiviral Pharmacology; 1993 - 95
- Nursing Pharmacology (200 students): Endocrine Pharmacology; 1994 - 2007
- Nursing Pharmacology (200 students): Antiviral Pharmacology; 1993 - 95
- Methods in Molecular Biology (graduate level): DNA and PCR; 1993 - 2005
- Developmental Biology (graduate level): Early mouse development; 1993 - 2013
- Developmental Biology (undergraduate level): Early mouse development; 1994
- Genetics (graduate level): manipulation of the mouse genome; 1995
- Regulatory Biology: Endocrine system; hormone regulation; circadian rhythms; 1994 - 99
- Endocrine Systems, HBH633 (module, Grad. Program in Pharmacology); 2000 - 2008
- BNB 552; Techniques in Neurobiology; 2000
- Signal Transduction Course, HBH553; 2001 – 2005
- Signal Transduction Module of the Grad Cell Bio Course, 2002 –
- Undergraduate Pharmacology: Endocrine Pharmacology; 2002
- Techniques in mouse genetics, Neurobiology HBH561; 2003 - 2005
- Bioinformatics BSB515; 2003 – 2015
- Responsible Conduct of Research, GRD500; 2003 -
- Seminars on Grant Writing, Dept. Pharmacology, 2005 – 2006
- Clinical Experimental Research Course, 2012 (Metabolomic Determination of Beta-Cell Mass)
- Infectious disease small groups, medical students, 2020 –
- Basic Principles of Pharmacology small groups, 2020 -

INVITED PRESENTATIONS

1985 Expression of MHC-Class II antigen by gamma-interferon-treated thyrocytes and the induction of autoimmune thyroiditis. Merc Research Institute.

1987 Expression of the mouse En-1 gene in the developing embryo. Workshop on hematopoetic and stem cell development, NIHCD.

1987 Immunolocalization of mouse En-1 in early neurulation. Medical Biology Institute, San Diego.

1989 Using RACE instead of conventional cDNA cloning techniques. Symposium on PCR, Cetus Research Corporation.

1989 RACE: A rapid way to clone new genes. California Society of Pathologists.

1990 Nested RACE: Direct analysis of unknown genes. Symposium on PCR, UCSF.

1990 Establishment of the anterior-posterior axis in mammals: Clues from the expression pattern of a homeobox-containing gene. Workshop on VATER, an association of midline defects. NIHCD.

1990 Pattern formation in the vertebrate embryo. Dept. Oral Biology, UCSF

1991 Establishment of the anterior-posterior axis in mammals: Clues from the expression pattern of homeobox-containing genes. Dept Biology, Dartmouth University

1991 Establishment of the anterior-posterior axis in mammals: Clues from the expression pattern of homeobox-containing genes. Dept Pharmacology, SUNY Stony Brook

1991 Establishment of the anterior-posterior axis in mammals: Clues from the expression pattern of homeobox-containing genes. Dept Anatomy, University of Cincinatti

1992 Expression and function of Gbx-2 and Hox-2.9 in mouse embryogenesis. Radiobiology Institute, UCSF

1994 Gbx-2 and mouse embryogenesis. Samuel Lumdens Institute, Mt. Sinai University

1996 REST and neurogenesis. Dept Pediatrics, SUNY Stony Brook

1996 REST, a gene controlling neural and myogenesis. Forum on Inductive Mechanisms. Aso, Japan

1996 Molecular studies on vertebrate Phospholipase D genes. University of Gifu, Japan.

1996 Molecular studies on vertebrate Phospholipase D genes. Dept. Pharmacology, U. Pittsburgh

1996 Induction of neural gene expression by ectopic expression of dominant negative REST. Society for Neuroscience meeting. Washington, DC.

1996 Phospholipase D: New twists on an old story. INSERM Atelier meeting, Paris, France

1997 Recent progress on Phospholipase D genes. Onyx corporation, Richmond, CA

1997 Molecular studies on Phospholipase D. Dept. Cell and Molecular Biology, Duke University

1997 Vertebrate Phospholipase D genes. Dept. Pharmacology, University of Tennessee at Memphis

1997 Molecular studies on Phospholipase D. Dept. Biology, U. Delaware.

1997 Recent progress on Phospholipase D genes. Novartis corporation, Summit, New Jersey

1997 Vertebrate Phospholipase D genes. Dept. Pharmacology, Georgetown University

1997 Molecular Biology of Phospholipase D. IBC Conference on Signal Transduction Therapy, San Diego, CA

1997 Phospholipase D, a family of signal-transducing enzymes involved in secretion and reorganization of the actin cytoskeleton. Dept. Pharmacology, University of Illinois at Chicago

1997 The Molecular Biology of Phospholipase D. Keynote address, Southeastern Regional Lipid Conference, Cashiers, North Carolina

1997 Recent progress on Phospholipase D genes. Parke-Davis, Ann Arbor, Michigan

1998 Phospholipase D, a family of signal-transducing enzymes involved in secretion. Brookhaven Labs, New York

1998 Phospholipase D, a family of signal-transducing enzymes involved in secretion. Institute of Molecular Medicine and Genetics, Medical College of Georgia.

1998 Progress on cellular roles for Phospholipase D. FASEB, Vermont. (meeting co-chairman, platform speaker).

1998 Molecular Studies on Mammalian Phospholipase D. International Symposium on Phosphoinositide Signaling, Miyagi, Japan (platform speaker).

1998 Phospholipase D, a family of signal-transducing enzymes with roles in secretion. Dept. Cell Biology, New York University

1999 Phospholipase D and the regulation of vesicular trafficking. Onyx corporation, Richmond, CA

1999 Phospholipase D, a family of signal-transducing enzymes with roles in secretion. Cancer Center, UCSF.

1999 Molecular Studies on Phospholipase D. FASEB, Vermont. (platform speaker)

1999 Phospholipase D and the regulation of vesicular trafficking. Dept. Physiology, John Hopkins University.

1999 Cellular roles for Phospholipase D genes. Novartis corporation, Summit, New Jersey

1999 Regulation of PLD, a signal-transducing enzyme with roles in vesicular trafficking. Inserm, Nice, France

1999 Structure and regulation of Phospholipase D. Parke-Davis, Paris, France

1999 PLD, a family of signal-transducing enzymes with roles in vesicular trafficking. Dept. Pharm., Wayne State

1999 Phospholipase D, a family of signal-transducing enzymes with roles in vesicular trafficking. Dept. Biochemistry, St. Louis University.

2000 Phospholipase D, a family of signal-transducing enzymes with roles in vesicular trafficking. Dept. Biochemistry, Medical University of South Carolina.

2000 Phospholipase D, a family of signal-transducing enzymes with roles in vesicular trafficking. Dept. Cell Biology, Albert Einstein University

2000 Progress on cellular roles for Phospholipase D, ASPET Symposium, Boston (platform speaker).

2000 Physiological Studies on Phospholipase D. FASEB, Keystone, Co. (platform speaker)

2000 Progress on cellular roles for Phospholipase D, Novartis. New Jersey.

2000 Roles for Phospholipase D genes in exocytosis and cellular remodeling. Parke-Davis, Ann Arbor, Michigan

2001 Physiological Studies on Phospholipase D. FASEB, Tucson, Az. (platform speaker; conference organizer)

2002 Recent advances on Phospholipase D. Phospholipases FASEB meeting, Tucson, Az. (platform speaker)

2002 Phospholipase D, a family of signal-transducing enzymes with roles in vesicular trafficking, phototransduction, and cell morphology. LI Cancer Center seminar

2003 Studies on PLD, a regulator of exocytosis. Inserm, Strasbourg, France

2003 PLD1 regulation of Glut-4 fusion in the plasma membrane. FASEB, Snowmass, Co (platform speaker).

2003 Regulated exocytosis in neuroendocrine cells. Dept Medicine (Div. Endocrinology), SUNY Stony Brook

2003 Development of a novel system for RNAi, Invitrogen Corporation, San Diego, CA

2004 Functional roles for Phospholipase D. Phospholipases FASEB meeting, White Mountains, Georgia (session chair, platform speaker)

2004 Insulin-Stimulated glucose transporter trafficking and the lipid-signaling enzyme Phospholipase D, Garvin Institute, University of Sidney, Sidney, Australia

2005 Roles for Phospholipase D in Vesicular Trafficking and Plasma Membrane Fusion during Regulated Exocytosis. ASBMB meeting on Intracellular Lipid Dynamics, San Diego, CA

2005 Mito-PLD, a novel lipid-modifying enzyme that regulates mitochondrial fusion. Institute of Molecular Medicine and Genetics, Medical College of Georgia.

2005 Mito-PLD, a novel lipid-modifying enzyme that regulates mitochondrial fusion. Symposium on Molecular Medicine, St. Louis, Mo.

2005 Mito-PLD, a novel lipid-modifying enzyme that regulates mitochondrial fusion. Dept. Biochemistry, Medical University of South Carolina.

2006 From Protein A to RACE-PCR, Lipid Signaling, and Mitochondrial Fusion. Tskuba University, Tokyo, Japan.

2006 Roles for Phospholipase D in Membrane Fusion during Regulated Exocytosis and Mitochondrial Fusion. FASEB meeting, Saxton River, Vermont (platform speaker).

2007 A common role for a signaling lipid in Membrane Fusion during Regulated Exocytosis and Mitochondrial Fusion. Dept. of Pathology and Cell Biology, Columbia University.

2007 Lipid Signaling in Mitochondrial Fusion. "Frontiers in Mitochondrial Research: from Bioenergetics to Dynamics" workshop, the Castle of Bertinoro – Bertinoro, Forlì-Cesena, Italy

2007 Lipid signaling pathways regulating mitochondrial fusion, energetics, and translocation. Dept. Pharmacology, Emory University.

2008 Lipid signaling pathways regulating mitochondrial fusion, energetics, and translocation. FASEB meeting, Saxton River, Vermont (platform speaker).

2008 Mitochondrial Fusion. Chonnam University, Gwangju, South Korea

2008 Phospholipase D superfamily members and their biological roles. Pohang University, POSTECH Biotech Center, South Korea

2008 Medical Scientist Training Programs. Ajou University, South Korea

2008 Lipid signaling pathways regulating mitochondrial fusion and energetics. Dept. Cell Biology, New York University

2009 Lipid signaling pathways regulating mitochondrial fusion and energetics. Keystone Symposium

2009 Medical Scientist Training Programs. Panel Discussion, Columbia University.

2010 Phospholipase D1 and lipid signaling roles in platelet activation and thrombus formation. Columbia University, Dept. Pharmacology.

2010 Unexpected role for the lipid signaling enzyme Phospholipase D1 (PLD1) in integrin activation and pathological angiogenesis. LI Cancer Center seminar

2010 New roles for Phospholipase D family members in platelet activation, secretion, and fertility. Institute of Molecular Medicine and Genetics, Medical College of Georgia.

2010 Roles for Phospholipase D family members in insulin secretion and Golgi complex organization, thrombosis, and RNAi activation during gametogenesis. Albert Einstein College of Medicine Diabetes Research Center

2010 Phospholipase D (PLD) regulation and targets. Acta Physiologica International Symposium on G-protein Coupled Receptors, Helsinki, Finland.

2010 A mitochondria surface lipid-signaling pathway links mitochondrial morphology to the spermatogenic piRNA pathway. Dept of Medicine, Boston University School of Medicine.

2011 Medical Scientist Training Programs. Chonnam University, Gwangju, South Korea

2011 JDRF Educational Conference, New York; Keynote speaker

2011 MRI Determination of Beta-Cell Mass. Endocrine Division, Stony Brook University

2011 Metabolomic Determination of Beta-Cell Mass. Juvenile Diabetes Research Foundation workshop: targeting the pancreatic beta cell. New York City.

2012 A mitochondria surface lipid-signaling pathway links mitochondrial morphology to the spermatogenic piRNA pathway. Gordon Research Conference on Mitochondria and Chloroplasts

2012 Phospholipase D lipid signaling in cancer. FASEB conference, Steamboat, Co.

2012 Roles for Phospholipase D1 in both primary tumor growth and metastasis. Pathology Grand Rounds, Stony Brook University School of Medicine

2012 Phospholipase D signaling in diabetes and cardiovascular disease. New York Academy of Sciences Symposium on Inositol Phospholipid Signaling in Physiology and Disease, NYC.

2012 Roles for Phospholipase D1 in primary tumor growth and metastasis. Cancer Center, Stony Brook University

2012 Roles for Phospholipase D1 lipid signaling in primary tumor growth and metastasis. Columbia University

2013 Roles for Phospholipase D1 in primary tumor growth and metastasis. Breast Cancer Retreat, Stony Brook University and Carol Baldwin Cancer Center

2013 Integrating Basic research and translational medicine at Stony Brook University, SBU Faculty Senate

2014 Phospholipase D physiological roles. FASEB conference, Niagara Falls
2014 Mitochondrial Lipid Signaling. Keystone Conference, Santa Fe, NM.
2014 Phospholipase D lipid signaling in cancer and metabolism. FASEB conference, Steamboat, Co.
2014 Lipid signaling enzymes in cancer. Cancer Retreat, Stony Brook University
2014 Roles for Phospholipase D1 in breast cancer primary tumor growth and metastasis. Breast Cancer Retreat, Stony Brook University and Carol Baldwin Cancer Center
2014 Phospholipase D lipid signaling in cancer and metabolism. U. Iowa, Dept. of Pharmacology
2014 The Phospholipase D superfamily as therapeutic targets. SBU Dept. of Medicine Research Seminar Series
2014 Presentation to pre-med society, SBU
2015 Overview of the MSTP and training grant – Oral Biology Retreat, Stony Brook University
2015 What keeps me up at night? International Pharmacology Chairs meeting (ASPET), St. Kitts
2015 The Phospholipase D superfamily as therapeutic targets. SBU Biochemistry Society
2015 Phospholipase D lipid signaling in neurodegeneration, Inserm, Université de Strasbourg Strasbourg, France
2015 Phospholipase D in cancer. Cancer Retreat, Stony Brook University
2016 From Protein A to RACE-PCR, Lipid Signaling, and Therapeutics: An MSTP odyssey. Keynote speaker, Medical Scientist Training Program retreat, Penn State University
2016 Phospholipase D lipid signaling in cancer and metabolism. FASEB conference, Steamboat, Co.
2016 Phospholipase D lipid signaling in cancer and metabolism. Breast Cancer Retreat, Stony Brook University and Carol Baldwin Cancer Center
2017 The Phospholipase D superfamily as therapeutic targets. The Herbert and Nicole Wertheim Leadership in Healthcare and Medicine Lectureship, Department of Cell Biology and Pharmacology at Florida International University
2017 Intercultural Communications: Understanding our Global Campus, Stony Brook University
2017 Phospholipase D lipid signaling in cancer and metabolism. Augusta University.
2017 The Phospholipase D superfamily as therapeutic targets. Winthrop University
2017 Phospholipase D in cancer. Cancer Retreat, Stony Brook University
2017 Phospholipase D lipid signaling roles in breast cancer metastasis. Boston University, Dept. Pharmacology
2017 Potential role of an unusual Phospholipase D enzyme in breast cancer and development of inhibitors to target metastatic progression. Institute for Chemical Biology and Drug Discovery.
2017 Phospholipase D in cancer. 58th Advances in Biological Regulation Symposium, Bologna, Italy.
2018 Phospholipase D in cancer and fertility. East Carolina University, Greenville, NC.
2019 The Phospholipase D superfamily as therapeutic targets. LA Biomed / UCLA.
2020 The Phospholipase D superfamily as therapeutic targets. Scripps Florida
2021 Phospholipase D in breast cancer. SBU Cancer Center retreat
2021 Phospholipase D1 as a genetic cause of congenital heart disease. Department of Medicine, Stony Brook
2022 Synder Research Symposium, Phospholipase D activation of novel pathways in breast cancer, SBCC

Funding - all amounts shown represent direct costs per year

Past

March of Dimes, Basil O'Conner award (1993-1995). $30K / year.

NIH R01 HD29758 (1993-1998) Mammalian gastrulation genes. $154K in final year.

Parkinson's Disease Foundation (1998) Role of □-synuclein and PLD2 in Parkinson's Disease. $32K.

NIH program project HG01487 (1996-1999) New Sequencing technologies (P.I. Luryi, S.). $35K / year.

American Heart Association fellowship award to Dr. Olga Redina (7/97 - 6/99), $35K / year.

Onyx corporation (1996-1999) Mammalian Phospholipase D genes. $20K / year.

Novartis fellowship award to Dr. Juan Crosby (3/98 - 1/00), D.C. $57K (salary + $20K / year supplies).

NSF INT 9603360 Japan US Cooperative research award "Regulation and roles of signal transducing Phospholipase D in cellular responses" (4/97 - 4/00), $13K/ year.

Pharmaceutical support from Parke-Davis, 2000-, ~$37.5K

NIH R01 grant for FASEB conference on PLD in 2001, $8500.

American Heart Association fellowship award for Dr. Guangwei Du, 7/00-6/02. $35K / year.

Carol M. Baldwin Breast Cancer Research Award (7/2000-6/2002) Role of Phospholipase D in Taxane-induced apoptosis, $30K / year

NIH R01 GM54813 (1/1998-12/2001) Mammalian Phospholipase D genes, $170K, final year.

NIH R01 GM60452 (1/2000 - 12/ 2003) Role of Drosophila Phospholipase D in cellularization, $145K current year.

NSF INT-9909888 Japan US Cooperative research award "Physiological and cellular roles of signal transducing Phospholipase D", $13K/ year. 6/2000 - 5/2003.

American Diabetes Association, (1/2002 – 12/2004) "Regulation of Glut-4 translocation by the signaling enzyme phospholipase D1", $85K / year.

Catacosinos Award (7/02 – 6/03) "Novel endonuclease released from mitochondria during apoptosis" $30K

Invitrogen, The Inducible Mammalian RNAi "Gene Swap" system, $50K

Training Grant NRSA award to Gary Jenkins, 12/2001-12-2004, $60K per year.

NIH NRSA Fellowship for Mary LaLonde: Phospholipase D in Phototransduction", 8/03 - 8/04, $28K

NIH R01 DK64166 (5/2003 – 4/2008) "Role of Phospholipase D in Glut-4 translocation", $175K.

URM Supplement to DK64166 for Diandra Ayala, $38K, 2003-5.

United Mitochondrial Disease Foundation (7/1/2005-2/3/2008), "A novel gene involved in mitochondrial fusion", $62K

Scientist Development Award to Guangwei Du (Mentor; Du, PI), AHA, 1/2004-12-2007, $65K

URECA Summer Research Program award (mentored funding) to Krishna Sarma, 2007, $3.5K

NIH NRSA Fellowship for Mary Osisami: Phospholipase D in Glut-4 translocation ", 8/05 - 8/09, $38K

HHMI Undergraduate Research Scholar fellowship (mentored funding) to Krishna Sarma, 2007-8, $3.5K

URECA Summer Research Program award (mentored funding) to Alex Chamessian, 2008, $3.5K

Baldwin award for breast cancer (11/2007 – 5/2010), "A PLD inhibitor", $40K.

NRSA Postdoctoral award to Phyllis Tseng, 4/2007 – 3/2010, $50K.

Scientist Development Award to Ping Huang (Mentor; Huang, PI), AHA, 1/2007-12-2010, $65K

NIH R01 GM071520 (5/2006 – 4/2011) "Role of Phospholipase D1 in Regulated Exocytosis", $190K

NIH NRSA Fellowship for Oladapo Yeku: Role for PI4P5-Kinase in insulin secretion ", 8/08 - 8/11, $30K

Juvenile Diabetes Research Found. (9/08 – 8/12) "Imaging beta-cells using non-invasive 1H-MRS", $250K

Catacosinos Cancer Research Award. (5/11 – 11/12) "A novel neoangiogenesis and metastasis small molecule inhibitor", $30K.

NIH **NRSA** T32 Postdoctoral fellowship award to Dr. Devi Ekanayake, 10/11 – 9/14, $49K

NIH **NRSA** T32 Predoctoral fellowship to Julie-Ann Cavallo , 4/1/13 – 6/30/14, $30K

Carol Baldwin Breast Cancer Award (PI, Frohman) 7/1/13 – 6/30/15
Inhibition of PLD1 as a therapeutic approach in breast cancer $35,000

URECA summer fellowship for Danya Abedeen summer, 2015 $4,000

NIH R01 GM100109 (2012-2016, PI, Frohman) "Regulation of RNA processing on the mitochondrial surface by lipid signaling. $200,000

NIH GM100109-S1, Equipment supplement to purchase a Li-Cor Odyssey infrared imager, 2015 $53,175

NIH GM084251-S1 (PI, Frohman) 7/2016 - 6/2017. Supplement to purchase a Zeiss Axio observer. $83,000

NIH **NRSA** F31 Predoctoral fellowship to Rochelle Nelson 6/1/13 – 5/30/17 $40,000

NIH Reach award (PI, Frohman), 12/2015 – 12/2017 $100,000. Development of small molecule inhibitors for PLD6 as therapeutics for Triple Negative Breast Cancer

Carol Baldwin Breast Cancer Award (PI, Frohman), 7/2017 – 6/19 Role of PLD6 in growth of chemoresistant metastases, $35,000

NIH GM084251, (PI, Frohman), Lipid-signaling pathways regulating mitochondrial morphology, energetics, and movement. 7/1/09 - 4/30/20. $200,000

NIH R35 GM124846 (PI, Jia S.) Role: co-PI. Exploring Single-Molecule Biophotonics for High-Throughput, Multiplexed Super-Resolution Microscopy. 9/17-8/22. $125,000

F30 NRSA (PI, Alexander, T.) 2/16/18 – 11/15/22
NIH CA225058 $40,000
Exploiting Metabolic Changes in HCC for Cancer Therapy

R01 (PI, Kim, H.) Role: co-PI 4/2018 – 4/2023 5%
NIH CA218132 $25,000 0.6 months
Proteolytic control of DNA interstrand cross-link repair and genome integrity

<u>Active</u> – all amounts listed as direct costs / year

P01 (PI, Hannun, Y.) Role: Co-I (Leader, Project 3) 9/2019 - 8/2024
NIH CA097132 $150,000 0.12 months
Role of Sphingosine Kinase in P53 cancer biology

Stony Brook Cancer Center Support 8/25/2022- (annual) 10%
Role of PLD in breast cancer $132,000 1.2 months

<u>Training Grants (No Direct support for M. Frohman)</u>

Frohman, M.A., Principal Investigator (Program Director)
GM08444-24 7/1/03-6/30/25 20%
NIH-NIGMS $837,216
Medical Scientist Training Program (MSTP)

American Cancer Society Institutional Research Grant 1/1/22 – 12/31/24 10%
(PI & Program Director, Frohman) $160,000

Synder Scholars (Mark and Gloria Snyder) (PD, Frohman) 9/1/22 – 8/31/27 1%
(PD, Frohman) $40,000

Scholarships for graduate students conducting PhD research in Cancer Center labs

Students / Postdocs trained

Prior High School students: Shneha Shah, 2000-2001
Matthew Mifsud, 2000 – 2001
Sean Nair, 2014, Simons Fellow: 3rd place at the regional JSHS

Prior Undergraduate students:

Abby Leventner (Pharmacology major); 1999 – 2000; unknown

Virginie Olive (Pharmacology Exchange student, Nice, France), 2001; postdoc at UC Berkeley

Adam Cohen (SUNY Stony Brook) 2001-2; NYU medical school, Asst. Prof. Fox Chase Cancer Center

Suzette Briones (SUNY Stony Brook) 2002- Med School, SBU; Pediatrician, Montefiore Hospital

Rabia Razi (SUNY Stony Brook) 2001-2003; U. Rochester Med School, U. Chicago Internal Medicine Residency, currently internist in Los Angelos

Oladapo Yeku (Medgar Evers College), 2003-2004; Stony Brook U. MSTP, U. Pitt. Internal Medicine

Mayra Pastore (San Francisco State), 2005; currently U. Wisconsin Grad. Student

Yue Zeng (Stony Brook University), 2005-2006; grad stud. U. Indiana, postdoc. U. Michigan

Afoma Ndubuisi (University of Maryland), 2006; New York Medical College; Staten Island U. Hospital

Olusola Bamishigbin (University of Miami), 2006; Miss University of Miami; appeared on ABC Family's new show *Design By Jane* in her television debut in February, 2013

June Hu (Biochemistry), 2007 - 2009; in medical school, Ohio State University

Krishna Sarma (Biochemistry), 2007 – 09; currently, U. Nebraska MD-PhD program

-URICA summer fellowship, 2007

-HHMI fellowship, 2007-2008

- Huang, H., Gao, Q. Peng, X.X., Choi, S.-Y., Sarma, K., Ren, H., Morris, A.J., and Frohman, M.A. (2011) piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev. Cell*, 20:376-387.

- in MD-PhD Program, U. Nebraska, 2011; U Mich. Residency

Alexander Chamessian (Biochemistry), 2007- 09

-Excellence in Research Award, 2008, from Sigma Xi

-URICA summer fellowship, 2008 (declined)

-HHMI Summer fellowship, 2008

- in MD-PhD, MSTP program at Duke University, 2010 – 2019

-resident physician-scientist in the Physical Medicine and Rehabilitation program at Washington University School of Medicine in St. Louis and also a post-doctoral fellow in the lab of Dr. Robert Gereau in Washington University Pain Center. The broad focus of my research is the neurobiology of pain, with specific interests in early-onset idiopathic neuropathies and chronic musculoskeletal pain.

Alyssa Genna (Biochemistry), 2008 – 09

-Su, W., Yeku, O., Olepu, S., Genna, A., Park, J.-S., Ren, H., Du, G., Gelb, M.H., Morris, A.J., and Frohman, M.A. (2009) FIPI, a pan-Phospholipase D pharmacological inhibitor that alters cell spreading and inhibits chemotaxis. *Mol. Pharm.* 75: 437-46.

Priscilla Pennino (Biology, Honors College), 2010 – 2011.

Lilly (Hui) Cheng (Pharmacology), 2009 - 2011

Isaiah Schuster (SBU Pharmacology graduate, post-bac year), 2010-11; med school, Stony Brook University, 2011-15

Yorleny Vicioso (Biochemistry), 2011-2012 –postbac at NIH; currently in MSTP at Case Western U.

- MARCS Award, 2001-12
- Member of Phi Theta Kappa Honors Society,
- Recipient of SUNY Chancellor's Award 2010

Kelsey Castro (HHMI summer student, 2013)

Kelsey Truglio (Honors College, 2012, WISE program)

Anjali Kapur (Honors College, 2012 - 2014 , WISE program); Award for Academic Excellence, 2014 – 18, SBU Medical School

Sunny Cho (WISE program, 2013 -2014)

Jai-Qi (Jackie) Tang (Pharmacology Major, 2013 - 2015); SUNY Upstate, Pharmacology Grad. Program

Tenesha Boyd (Biochemistry major, 2013 - 2014)

Elizabeth Ha (Honors College, 2011 - 2015, WISE program, URECA summer fellowship, 2013); CAS scholarship for excellence, 2015; Columbia U. MSTP MD-PhD program; just for NRSA F30

Ha, E. and Frohman, M.A. (2014) Regulation of Mitochondrial Morphology by Lipids. *Biofactors*, 40(4):419-24.

Berke Gursoy (Cornell, summer 2015)

Jasper Rubin (ASPET SURF summer student, 2015; grad school at USC)

Obiora Egbo (IMSD program, 2014 - 15)

Danya Abedeen (WISE program, 2013 - 2016; URECA summer fellowship, 2015)

Krisann Marquart (Pharmacology, 2015 - 16)

Cassidy Guida (Rhodes College, Pharmacology SURF program, 2016)

Veevek Shah (Biochemistry, 2014 -; URECA Biology Alumni Research award, 2016; is technician at Mt. Sinai, now applying to med schools)

Serena Fleischman (Pharmacology, 2015 -17)

Andrew Chen (Pharmacology, 2016 -17)

Rafae Nasim (Carnegie Melon, SURF student, 2017; current NIH Post-bac program)

Daniela Barisano (WISE, Biochemistry, 2016 - 2019);

Oindrila Naha (WISE, Biochemistry, 2018-19);

Rushnan Islam (WISE, Pharmacology, 2018-19)

Roma Kidambi, BA-MD student, 2020-21

Anabelle Solano, 2020 – 2021

Sehaan Tarique, 2021 – 2022

Snowiya Multani, 2021 - 2022

Ellena Karapatakis, 2022 -

Prior Masters Students: Chelsea Byrd
Ya Wei

Prior Graduate Students:

Colley, W.C.; "Cloning and Characterization of Murine Phospholipase D genes"
(did post-doctoral training at St. Judes, Memphis, TN, Dept. Biochem.; deceased)

Sung, T.-C.; "Molecular Studies on Phospholipase D genes" (current: Staff Researcher, Salk Institute)

Zhang, Yue; "Regulation of Phospholipase D by Rho and Protein Kinase C"
(Research Asst. Professor, Stony Brook, Dept. Microbiology)

Yun, S.; "Role of Phospholipase D in Drosophila cellularization" (Sr. Manager, Clinical Operations at Eisai Medical Research)

Wu, Si; "Role of PIP2-interaction in PLD1 localization" [Masters degree] (Currently enrolled in graduate school in computers)

Danxia Ke; "Investigation of K4, a distantly-related PLD superfamily member" [Masters degree] (currently enrolled in post-graduate studies in applied math)

Nianzhou Xiao: "PrxII, a signal terminator for Phospholipase D1" (Currently engaged in medical residency at Cleveland Clinic)

Seok-Yong Choi:"Mito-PLD, a novel PLD superfamily member that regulates mitochondrial fusion" (Current: associate professor in the department of Biomedical Science at Chonnam University Medical School, Gwangju, South Korea)

Liz Scotto-Lavino: "Identification of Novel Interaction Sites between Myosin Phosphatase (MP) subunitsPP1c β and MYPT1". Microbiologist at Altaire Pharmaceuticals

Wenjuan Su (Pharmacology), 2004 – 2009. Postdoctoral fellowship, NYU.

Mary Osisami (Genetics); "Exploring a Role for PLD3 during *in vitro* myogenesis" (current: post-doc, University of Michigan)

Winnie Huang (Pharmacology), 2006 –2011. Postdoc, Brown University

Akua Bonsra (Co-mentor with Dr. Guangwei Du), 2005 – 2010. Regulatory Affairs Specialist at Technical Resources International, Inc

Wahida Ali (Pharmacology), 2009 – 2013. Postdoc, University of Michigan; Faculty member, Columbus State college

Pheobee (Chai-Ying) Yang, 2010 – 2013. Returned to China; currently on maternity leave.

Julie Cavallo, 2012 – 2016, technician position at Sloan Kettering; Postdoctoral Fellow, Mount Sinai School of Medicine

Rochelle Nelson, 2011 – 2017, "Effects of Phospholipase D2 on Cardiovascular Function and Disease" postdoctoral position at Albert Einstein; Assistant Professor; Department of Biological Sciences and Geology; Queensborough Community College

Chris Salazar, 2017 – 2021. "Ablation of PLD1 and 2 Increases Epithelial Cell Differentiation in the MMTV-PyMT Mouse Model". Biotech.

Prior M.D. Ph.D. (MSTP) Students:

Margaret LaLonde (Ph.D., Genetics), 2001-5. Fast-Track, Academic Research-oriented, Psychiatry Residency at Mt. Sinai (Class of 2012); Asst. Prof. Mt Sinai Dept. Psychiatry.

Oladapo Yeku (Ph.D., Pharmacology), 2006 – 2010 (Heme-Onc fellowship & tumor immunotherapy program at Sloan Kettering)

Eric Roth (Ph.D., Pharmacology), 2016 – 2019. Phospholipase D1 and D2 in Breast Cancer Growth and Metastasis. Residency (PSTP) UCSD.

Prior Postdocs:

Dr. Olga Redina; 1996 - 1999. Present: Professor, U. Siberisk, Russia

Dr. Yue Zhang; 1999-2000. Present, Research Asst. Prof., Dept. Microbiology, Stony Brook University

Dr. Juan Crosby, 1997 - 2000. Did postdoc UMNJ, lab of Dr. Marjorie Brandeis (lost contact)

Dr. Guangwei Du; 1999 – 2001. Present, Associate Professor, U. Texas at Houston, Department of Integrative Biology and Pharmacology

Dr. Hilde Janssens (Ph.D., U. Basel, Switzerland); 2000 – 2002; Flymine, Cambridge U., UK

Dr. Gary Jenkins (Ph.D., Duke University); 2001 – 2005

Dr. Xiao Yu (Ph.D. Institute of Neurosciences, Shanghai, China); 2004 – 2005; Present, postdoctoral fellow, Indiana University

Dr. Seok-Yong Choi (Ph.D. Stony Brook University); 2005-6; Present, Associate Professor, Dept of Biomedical Science, Chonnam University Medical School, Gwangju, South Korea

Dr. Ping Huang (M.D., Norman Bethune U, China; Ph.D., U. Hong Kong); 2000-7; Instructor, Harvard; Assistant Professor, Department of Pathology, Medical School of Jilin University

Dr. Phyllis Tseng (Ph.D. U. Hong Kong, 2008); 2008 – 09; last known position, postdoctoral fellow, University of Hong Kong

Dr. Winnie Huang, (Ph.D., Stony Brook University, 2010); 2010 – 2011; postdoc, Brown University

Dr. Qin Chen (Ph.D., Shanghai Institute of Materia Medica, 2008); 2009 –2012; Associate Professor, Shanghai Jiao Tong University

Dr. Qun Gao (Ph.D. New York Medical School, 2008); 2009 – 2012. Postdoc, Dept. Anesthesiology, SBU

Xiaoxue Peng (Ph.D. Shanghai U., 2008); 2009 –2013. Postdoc, Dept. Neurology, SBU

Devi Ekanayake (Ph.D., U. Kansas, 2010); 2011 – 2014; biotechnology position with Genesys.

Yi Zhang (Ph.D., 2011); 2012 –16

Prior Research Asst. Professor:

Dr. Guangwei Du, (Ph.D, Peking Union Medical College); 2001 – 2008. Present, Associate Professor, U. Texas at Houston.

Dr. Finly Philip (Ph.D. 2007); 2014 – 2018.

Present undergraduate students:

Jackson Chan, 2022 –

Publications h-index = 88 (Google Scholar)

1. Frohman M.A., Frohman, L.A., Goldman, J., Goldman, M. (1979). Use of Protein A-containing Staphylococcal aureus as an immunoadsorbant in radioimmunoassays to separate antibody-bound from free antigen. *Journal of Laboratory Clinical Medicine* 93, 614-621.

2. Daddona, P.E., Frohman, M.A. and Kelley, W.N. (1979). Radioimmunochemical quantitation of human adenosine deaminase. *Journal of Clinical Investigation* 64, 798-803.

3. Daddona, P.E., Frohman, M.A. and Kelley, W.N. (1980). Human adenosine deaminase and its binding protein in normal and adenosine deaminase deficient fibroblast cell strains. *Journal of Biological Chemistry* 255, 5681-5687.

4. Daddona, P.E., Frohman, M.A. and Kelley, W.N. (1980). Radioimmunochemical analysis of human erythrocyte adenosine deaminase. In: *Purine Metabolism in Man*, Vol. III, Part B. A. Rapado, R.W.E. Watts and C.H.M.M. DeBruyn. (Plenum Publishing Co.).

5. Daddona, P.E., Frohman, M.A. and Kelley, W.N. (1981). Radioimmunoassay of human adenosine deaminase. In: *Methods in Enzymology*, Vol. 74, 351-378.

6. Frohman, L.A., Berilowitz, M., Gonzalez, C., Barowsky, H., Rao, R., Lim, V.S., Frohman, M.A. and Thominet, J.L. (1981). Studies of dopaminergic mechanisms in hyperprolactinemic states. In: *Serono Clinical Symposia on Reproduction*, Vol. 2. P. Crosignani and B. Rubin, eds. (Academic Press, London). 39-52.

7. Cowing, C., Frohman, M.A. and Chapdelaine, J.A. (l984). Activation of T cells requires asialo-Ia on antigen-presenting cells. In: *Regulation of the Immune System*. E. Sercarz, H. Cantor and L. Chess, eds. (Alan R. Liss, New York).

8. Francfort, J.W., Barker, C.F., Kimura, H., Silvers, W.K., Frohman, M.A. and Naji, A. (l985). Increased incidence of Ia antigen-binding T lymphocytes in the spontaneously diabetic BB rat. *Journal of Immunology* 134, 1577-1582.

9. Frohman, M. A. and Cowing, C. (1985). Presentation of antigen by B cells: functional dependence on radiation dose, interleukins, cellular activation and differential glycosylation. *Journal of Immunology* 134, 2269-2275.

10. Wolgemuth, D.J., Viviano, C.M., Gizang-Ginsberg, E., Frohman, M.A., Joyner, A.L. and Martin, G.R. (l987). Differential expression of the mouse homeo box-containing gene Hox-1.4 during male germ cell differentiation and embryonic development. *Proceedings of the National Academy of Science, USA* 84, 5813-5817.

11. Martin, G.R. and Frohman, M.A. (1988). The mouse *En-1* gene encodes a nuclear protein expressed in the nervous system of the developing embryo. In: *Cell to Cell Signals in Mammalian Development*. NATO ASI Series. S. de Laat, J. G. Bluemink and C.C. Mummery, eds. (Springer-Verlag). 43-52.

12. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Science, USA* 85:8998-9002.

13. Frohman, M.A., Downs, T.R., Chomczynski, P., and Frohman, L.A. (1989). Cloning and characterization of mouse growth hormone-releasing hormone (GRH) cDNA: Increased GRH mRNA levels in the growth hormone-deficient lit/lit mouse. *Molecular Endocrinology* 3:1529-1536.

14. Frohman, M.A. and Martin, G.R. (1989). Rapid amplification of cDNA ends using nested primers. *Technique* 1:165-170.

15. Frohman, M.A. and Martin, G.R. (1989). Cut, paste, and save: New approaches to altering specific genes in mice. *Cell* 56, 145-147.

16. Frohman, M.A. (1989). Creating full-length cDNAs from small fragments of genes: Amplification of rare transcripts using a single gene-specific oligonucleotide primer. In: *PCR Protocols and Applications: A Laboratory Manual*., ed. Innis, M., Gelfand, D., Sninsky, J., and White, T. 28-38.

17. Frohman, M.A. (1989) Rapid amplification of cDNA ends using nested primers, a simple and new technique for cDNA cloning. The Scientist.

18. Frohman, M.A. (1990). Detection of homologous recombinants. In: *PCR Protocols and Applications: A Laboratory Manual*., ed. Innis, M., Gelfand, D., Sninsky, J., and White, T. 228-236.

19. Frohman, L.A., Downs, T.R., Chomczynski, P., and Frohman, M.A. (1990). Growth hormone-releasing hormone: Structure, gene expression, and molecular heterogeneity. *Acta Pediatrics Scandinavia* 367:81-86.

20. Frohman, M.A., Boyle, M., and Martin, G.R. (1990). Isolation of the mouse *Hox-2.9* gene: analysis of embryonic expression suggests that positional information along the anterior-posterior axis is specified by mesoderm. *Development* 110:589-608.

21. Mizobuchi, M., Frohman, M.A., Downs, T.R., Frohman, L.A. (1990) Tissue-specific transcription initiation and growth hormone (GH) feedback regulation of mouse and rat GH-releasing hormone (GRH) mRNA in hypothalamus and placenta. *Molecular Endocrinology* ***5***:476-84.

22. Frohman, M.A. (1990). Rapid Amplification of cDNA Ends (RACE): User Friendly cDNA Cloning. *Amplifications* 5:11-15.

23. Frohman, M.A., Francfort, J.W., and Cowing, C. (1991) T-dependent destruction of thyroid isografts exposed to IFN-gamma. *Journal of Immunology* 146:2227-2234.

24. Frohman, M.A. and Martin, G.R. (1992). Isolation and analysis of embryonic expression of Hox-4.9, a member of the murine labial-like gene family. *Mechanisms in Development* 38:55-67.

25. Frohman, M.A., Martin, G.R., and Barsh, G.B. (1993) Altered rhombomere-specific gene expression and hyoid bone differentiation in the mouse segmentation mutant, kreisler (kr). *Development* 117:925-936.

26. Frohman, M.A., Dickinson, M.E., Hogan, B.L.M., and Martin, G.R. (1993) Localization of two new and related homeobox-containing genes to chromosomes 1 and 5, near the phenotypically similar mutant loci dominant hemimelia (Dh) and hemimelic extra-toes (Hx). *Mouse Genome* 91:323-325.

27. Bulfone, Puelles, L., Porteus, M.H., Frohman, M.A., Martin, G.R., and Rubenstein, J.R.L. (1993) Spatially-restricted expression of Dlx-1, Dlx-2, Gst-2 and Wnt-3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. *Journal of Neuroscience* 13:3155-3172.

28. Frohman, M.A. (1994) On Beyond Classic RACE. *PCR Methods and Applications* 4:S40-58.

29. Carroll, PM, Tsirka, SE, Richards, WG, Frohman, MA and Strickland, S (1994) The mouse tissue plasminogen activator gene 5' flanking region directs appropriate expression in development and a seizure-enhanced response in the CNS. *Development* 120: 3173-83.

30. Frohman, M.A. (1994) Thermal RACE (Rapid Amplification of cDNA Ends) as an alternative to conventional cDNA cloning. *Methods in Enzymology* 218:340-356.

31. Frohman, M.A. (1994) Cloning PCR products: Strategies and tactics. In: *PCR. The Polymerase Chain Reaction. Methods in Molecular Biology Series.* , ed. Mullis, K.B., Ferre, F., Gibbs, R.A. 14 - 37.

32. Chong, J.A., Tapia-Ramírez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuller, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995) REST: A Mammalian Silencer Protein that Restricts Sodium Channel Gene Expression to Neurons. *Cell* 80:949-957

33. Hammond, S.M., Altshuller, Y.M., Sung, T.-C., Rudge, S.A., Rose, K., Engebrecht, J., Morris, A.J, and Frohman, M.A. (1995) Human ARF-activated phosphatidylcholine-specific Phospholipase D Defines a New and Highly Conserved Gene Family. *Journal of Biological Chemistry* 270: 29640-43.

34. Rose, K., Rudge, S., Frohman, M., Morris, A., and Engebrecht, J. (1995). Phospholipase D signalling is essential for meiosis. *Proc Natl Acad Sci U.S.A.* 92:12151-12155.

35. Frohman, M.A. (1995) Rapid Amplification of cDNA Ends. In *PCR Primer, A Laboratory Manual.* (ed., Dieffenbach, C.W. and Dveksler, G.S.), CHSL Press, Cold Spring Harbor. 381-409.

36. Altshuller, Y.M., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Frohman, M.A. (1996) *Gcm1*, a mammalian homolog of Drosophila *Glial Cells Missing. FEBS Letters* 393:201-204.

37. Morris, A.J., Engebrecht, J., and Frohman, M.A. (1996) Structure and Regulation of Phospholipase D. *Trends in Pharmacological Sciences* 5:182-185.

38. Frohman, M.A. and Morris, A.J. (1996) Phospholipid Signaling: Rho is only ARF the story. *Current Biology* 6:945-947.

39. Frohman, M. A. and Morris, A. J. (1996). Molecular studies on phospholipase D genes: New twists on an old story. In *Lipases and transduction: Molecular tools* (F. Souyri, ed). Paris: Inserm. 87:250-258.

40. Zhang, Y. and Frohman, M.A. (1997) Using Rapid Amplification of cDNA Ends (RACE) to obtain full length cDNAs. In *cDNA Library Protocols* (Cowell, I.G. and Austin, C.A. ed.) Humana, Totowa. 69:61-87.

41. Hammond, S. M., Jenco, J. M., Nakashima, S., Cadwallader, K., Cook, S., Nozawa, Y., Frohman, M. A., and Morris, A. J. (1997) Characterization of two alternately spliced forms of Phospholipase D1. Activation of the purified enzymes by PIP_2, ARF, and Rho family monomeric G-proteins and protein kinase C- □□□*Journal of Biological Chemistry* 272:3860-3868.

42. Chen, Y-G, Siddhanta, A., Austin, C.D., Hammond, S.M., Sung, T.-C., Frohman, M.A., Morris, A.J and Shields, D. (1997) Phospholipase D Stimulates Release of Nascent Secretory Vesicles from the *trans*-Golgi Network. *Journal of Cell Biology* 138:495-504.

43. Colley, W.C., Sung, T.-C., Roll, R., Jenco, J., Hammond, S.M., Altshuller, Y.M., Bar-Sagi, D., Morris, A.J., and Frohman, M.A. (1997) Phospholipase D2, a distinct Phospholipase D isoform with novel regulatory properties that provokes cytoskeletal reorganization. *Current Biology* 7:191-201.

44. Morris, A.J., Frohman, M.A., and Engebrecht, J. (1997) Measurement of Phospholipase D activity. *Analytical Biochemistry* 252:1-9.

45. Liang, J.O., Sung, T.-C., Morris, A.J., Frohman, M.A. and Kornfeld, S. (1997) Different domains of mammalian ADP-ribosylation factor 1 mediate interaction with selected target proteins. *Journal of Biological Chemistry*, 272:33001-8.

46. Sung, T.-C., Roper, K., Zhang, Y., Rudge, S., Temel, R., Hammond, S.M., Morris, A.J., Moss, B., Engebrecht, J., and Frohman, M.A. (1997) Mutagenesis of Phospholipase D defines a superfamily including a trans-Golgi viral protein required for poxvirus pathogenicity. *EMBO Journal* 16:4519-4530.

47. Morris, A.J., Hammond, S.M., Colley, W.C., Sung, T.-C., Jenco, J.M., Sciorra, V.A., Rudge, S.A., and Frohman, M.A. (1997) Regulation and functions of Phospholipase D. *Biochemical Society Transactions: The Cell Biology of Lipid Signalling* 25:1151-1157.

48. Colley, W.C., Altshuller, Y.M., Zhang, Y., Sue-Ling, C.K., Copeland, N.G., Gilbert, D.J., Jenkins, N.J., Branch, K.D., Tsirka, S.E., Bollag, R.J., Bollag, W.B., and Frohman, M.A. (1997) Cloning and expression analysis of murine PLD1. *Biochemical Journal* 326:745-753.

49. Luo, J.-Q., Liu, X., Hammond, S.M., Colley, W.C, Feig, L.A., Frohman, M.A., Morris, A.J., and Foster, D.A. (1997) Ral A interacts directly with the ARF-responsive PIP_2-dependent Phospholipase D1. *BBRC* 235:854-859.

50. Zhang, Y. and Frohman, M.A. (1998) Cloning cDNA ends using RACE. In: *Methods in Molecular Medicine* (Reischl, U. ed.) Humana, Totowa. 13:81-105.

51. Frohman, M.A. (1998) Molecular Biology of Phospholipase D. In: *Cell Signal Transduction* (Minden, S.A. and Guttry, P. ed.) IBC Library Series, Southborough. pp 65-70.

52. Nakashima, S., Ohguchi, K., Frohman, M.A., and Nozawa, Y. (1998) Increased mRNA expression of phospholipase D (PLD) isozymes during granulocytic differentiation of HL60 cells. *Biochim. Biophys. Acta*, 1389:173-177.

53. Redina, O. and Frohman, M.A. (1998) Genomic analysis of murine *Phospholipase D1* and comparison to *Phospholipase D2* reveals an unusual difference in gene size. *Gene* 222:53-60.

54. Redina, O. and Frohman, M.A. (1998) Organization and alternate splicing of the murine Phospholipase D2 gene. *Biochemical Journal* 331:845-851.

55. Slaaby, R., Jensen, T., Hansen, H.S., Frohman, M.A., and Seedorf, K. (1998) PLD2 complexes with the EGF receptor and undergoes tyrosine phosphorylation at a single site upon agonist stimulation. *Journal of Biological Chemistry* 273: 33722-33727.

56. Frohman, M.A., Engebrecht, J., and Morris, A.J. (1999) Phospholipase D structure and regulation. *Chemistry and Physics of Lipids*. 98: (1-2) 127-140.

57. Frohman, M.A., Sung, T.-C., and Morris, A.J. (1999) Phospholipse D structure and regulation. In: *Phospholipase D* (Daniel, L.W. ed.) *Biochimica Biophysica Acta - Lipids and Lipid metabolism*, 1439:175-186.

58. Sung, T.-C., Altshuller, Y.M., Morris, A.J. and Frohman, M.A. (1999) Molecular Analysis of Mammalian Phospholipase D2. *Journal of Biological Chemistry* 274:494-502.

59. Sung, T.-C., Zhang, Y., Morris, A.J. and Frohman, M.A. (1999) Structural Analysis of Human Phospholipase D1. *Journal of Biological Chemistry*, 274: 3659-3666.

60. Zhang, Y. and Frohman, M.A. (1999) Rapid Amplification of cDNA Ends (RACE) to obtain full length cDNAs. In: *The Nucleic Acids Protocols Handbook* (Walker, J. ed.) Humana, Totowa. 81-106.

61. Rizzo, M.A., Shome, K., Vasudevan, C., Stolz, D.B., Sung, T.-C., Frohman, M.A., Watkins, S.C., and Romero, G. (1999) Phospholipase D and its product, phosphatidic acid, mediate agonist-dependent Raf-1 translocation to the plasma membrane and the activation of the MAP kinase pathway. *Journal of Biological Chemistry*, 274: 1131-1139.

62. Yamazaki, M., Zhang, Y., Watanabe, H., Yokozeki, T., Ohno, S., Kaibuchi, K., Shibata, H., Mukai, H., Ono, Y., Frohman, M.A., and Kanaho, Y. (1999) Identification of the RhoA-binding domain of Phospholipase D1. *Journal of Biological Chemistry*, 274:6035-6038.

63. Sciorra, V.A, Rudge, S., Prestwich, G.D., Frohman, M.A, Engebrecht, J., and Morris, A.J. (1999) Identification of a phosphoinositide binding motif that mediates activation of mammalian and yeast Phospholipase D isoenzymes. *EMBO J.*, 18: 5911-5921.

64. Zhang, Y, Altshuller, Y.A., Hammond, S.A., Hayes, F., Morris, A.J., and Frohman, M.A. (1999) Loss of Receptor Regulation by a Phospholipase D1 mutant unresponsive to Protein Kinase C. *EMBO J.*, 18:6339-6348.

65. Honda, A., Nogami, M., Yokozeki, T., Yamazaki, M., Nakamura, H., Watanabe, H., Kawamoto, K., Nakayama, K., Morris, A.J., Frohman, M.A., and Kanaho, Y. (1999) Phosphatidylinositol 4-phosphate 5-kinase□ is a Downstream Effector of the Small G Protein ARF6 in Membrane Ruffle Formation. *Cell*, 99:521–532.

66. Sciorra, V.A., Frohman, M.A., and Morris, A.J. (2000) Regulation of Phospholipase D signaling by phosphoinositides. In: The Biology of Phosphoinositides (Cockcroft, S. ed.) Series: Frontiers in Molecular Biology, Oxford University Press. pp 269-297.

67. Wu, Y.P., Lu, W., Sung, T-C., Frohman, M.A., Milev, P., Bugge, T.H., Degen, J.L.,Margolis, R.U., and Tsirka, S.E. (2000) The tPA/plasmin extracellular proteolytic system contributes to hippocampal mossy fiber reorganization through a novel proteoglycan substrate. *J. Cell Biol.*, 148:1295-304. PMCID: PMC2174310

68. Frohman, M.A., Kanaho, Y., Zhang, Y., and Morris, A.J. (2000) Regulation of Phospholipase D1 activity by Rho GTPases. In : Regulators and effectors of small GTPases. Methods in Enzymology (Hall, A. Editor). 325:177-189.

69. Slaaby, R., Du, G., Altshuller, Y. M., Frohman, M.A., and Seedorf, K. (2000) Insulin-induced phospholipase D1 and phospholipase D2 activity in human embryonic kidney-293 cells mediated by the PLC□ and PKC□ signaling cascade. *Biochemical Journal,* 351:613-619. PMCID: PMC1221400

70. Lee, C., Kim, S.R., Chung, J.-K., Frohman, M.A., Kilimann, M.W., Rhee, S.G. (2000) Inhibition of Phospholipase D by Amphiphysins. *Journal of Biological Chemistry*, 275: 18751-18758.

71. Frohman, M.A. and Tsirka, S.E. (2000) Polymerase Chain reaction: specialized applications. In: Nature Encyclopedia of Life Sciences (Robertson, S., Editor). Macmillan, London.

72. Zhang, Y., Redina, O., Altshuller, Y.M., Yamazaki, M., Ramos, J., Chneiweis, H., Kanaho, Y., and Frohman, M.A. (2000) Regulation of PLD1 and PLD2 expression levels by Pea-15, a novel PLD-interacting protein. *Journal of Biological Chemistry*, 275: 35224-35232.

73. Du, G., Altshuller, Y.M., Kim., Y., Han, J.M., Ryu, S.H., Morris, A.J., and Frohman, M.A. (2000) Dual requirement for Rho and protein kinase C in direct activation of phospholipase D1 through G-protein coupled receptor signaling. *Molecular Biology of the Cell*, 11:4359-4368. PMCID: PMC15078

74. Park, J.B., Kim, J.H., Kim, J.H., Kim, J, Yoo, J.-S.,Suh, P.-G., Du, G., Frohman, M.A., and Ryu, S.H. (2000) Phospholipase D2 is localized in sarcolemmal membrane and inhibited by □-actinin through direct interaction. *J. Biological Chemistry*, 275: 21295-21301.

75. Du., G., Frohman, M.A. and Morris, A.J. (2001) Phospholipase D: Isoforms and Function. In: *Annual Review of Pharmacology and Toxicology* (Cho, A.K. ed.). Volume 41. In press.

76. Vitale, N., Caumont-Primus, A.-S., Chasserot-Golaz, S., Du, G., Wu, S., Sciorra, V.A., Morris, A.J., Frohman, M.A. and Bader, M.-F. (2001) Phospholipase D1: a key factor for the exocytotic machinery in neuroendocrine cells. *EMBO J.*, 20:2424-34. PMCID: PMC125248

77. Freyberg, Z., Siddhanta, A., Sweeney, D., Bourgoin, S., Frohman, M.A., and Shields, D. (2001) Intracellular Localization of Phospholipase D1 in Mammalian Cells. *Molecular Biology of the Cell*, 12:943-955. PMCID: PMC32278

78. Frohman, M.A. and Engebrecht, J.E. Phospholipase D in cell signaling. (2001) In: *Molecular Biology of the Cell*. Eds. Purton, M. Chapter 15.

79. Zhao, D., Frohman, M.A., and Blusztajn, J.K. (2001) Generation of choline for acetylcholine synthesis by Phospholipase D isoforms. *BMC Neuroscience* 2:16-24. PMCID: PMC60648

80. Humeau, Y., Vitale, N., Chasserot-Golaz, S., Dupont, J.-L., Du, G., Frohman, M.A., Bader, M.-F. and Poulain, B. (2001) A role for phospholipase D1 in neurotransmitter release. *PNAS* 98:15300-5. PMCID: PMC65024

81. Denmat-Ouisse, L.-A., Phebidias, C., Honkavaara, P., Robin, P., Frohman, M.A., and Raymond, M.-N. (2001) Regulation of constitutive protein transit by PLDs in HT29-cl19A cells. *J. Biol. Chem.* 276:48840-6.

82. Du, G., Morris, A.J., Sciorra, V.A., and Frohman, M.A. (2002) G-protein coupled receptor regulation of Phospholipase D. In: G-protein Pathways. *Methods in Enzymology* (Hildebrandt, J.D. and Iyengar, R., Editor). 345:265-274.

83. Yamazaki, M., Miyazaki, H., Watanabe, H., Maehama, T., Frohman, M.A., and Kanaho, Y. (2002) Phosphatidylinositol 4-phosphate 5-kinase couples ROCK to neurite remodeling. *J. Biol. Chem.*, 277:17226-17230.

84. Choi., W.S., Kim, Y.M., Combs, C., Frohman, M.A., and Beaven, M.A. (2002) Phospholipase D1 and 2 Regulate Different Phases of Exocytosis in Mast Cells. *J. Immunology.* 168:5682-9.

85. Vitale, N., Chasserot-Golaz, S., Bailly, Y., Morinaga, N., Frohman, and Bader, M.-F. (2002) Calcium-regulated exocytosis of dense core vesicles requires the activation of ARF6 by ARNO at the plasma membrane. *J. Cell Biol.*159:79-89. PMCID: PMC2173505

86. Fang, Y., Park, I.-H., Wu, A.-L., Du, G., Huang, P., Frohman, M.A., Walker, S.J., Brown, H.A., Chen, J. (2003) PLD1 regulates mTOR signaling and mediates Cdc42 activation of S6K1. *Current Biology* 13:2037-44.

87. Du, G., Altshuller, Y.M., Vitale, N., Huang, P., Chasserot-Golaz, S.,Morris, A.J., Bader, M.F., and Frohman, M.A. (2003) Regulation of Phospholipase D subcellular cycling through coordination of multiple membrane association motifs. *J. Cell Biol* 62:305-15.

88. Huang, P. and Frohman, M.A. (2003) The role of phospholipase D in Glut-4 translocation. *Diabetes Metabolism Research and Reviews*. 19:456-63.

89. Frohman, M.A. (2003) Rapid Amplification of cDNA Ends. In *PCR Primer, A Laboratory Manual (2nd edition)*. (ed., Dieffenbach, C.W. and Dveksler, G.S.), CHSL Press, Cold Spring Harbor. In press.

90. Ballou, L.M., Jiang, Y.-P., Du, G., Frohman, M.A., and Lin, R.Z. (2003) Ca2+- and Phospholipase D-dependent and -independent Pathways Activate mTOR Signaling. *FASEB Letters,* 550:51-6.

91. Zhang, Y., Huang, P., Du, G., Kanaho, Y., Frohman, M.A., and Tsirka, S.E. (2004) Increased Expression of Two Phospholipase D Isoforms during Experimentally Induced Hippocampal Mossy Fiber Outgrowth. *Glia*, 46:74-83.

92. Watanabe, H., Yamazaki, M., Miyazki, H., Itoh, K.-I., Sasaki, T., Maehama, T., Frohman, M.A., and Kanaho, Y. (2004) Phospholipase D2 functions as a downstream signaling molecule of MAP kinase pathway in L1-stimulated neurite outgrowth of cerebellar granule neurons. *J. Neurochem.*, 1:142-51.

93. LaLonde, M. and Frohman, M.A. (2004) Phospholipase D. In *Encyclopedia of Biological Chemistry*. (eds. Lennarz, W. and Lane, M.D.), Academic Press / Elsevier, New York, 3:306-313.

94. Zheng, X.-L., Gui, Y., Du, G., Frohman, M.A., and Peng, D.-Q. (2004) Calphostin-C induction of vascular smooth muscle cell apoptosis proceeds through phospholipase D and microtubule inhibition. *J. Biol. Chem.*, 279:7112-7118.

95. Du, G., Huang, P., Liang, B.T. and. Frohman, M.A. (2004) Phospholipase D2 localizes to the plasma membrane and regulates Angiotensin II receptor endocytosis. *Molecular Biology of the Cell,* 15:1024-30. PMCID: PMC363061

96. Hughes, W. E., Elgundi, Z., Huang, P., Frohman, M.A., and Biden, T.J. (2004) Phospholipase D1 regulates secretagogue-stimulated insulin release in pancreatic □-cells. *J. Biol. Chem.* 279:27534-41.

97. Watanabe, H., Yokozeki, T., Yamazaki, M., Miyazaki, H., Sasaki, T., Maehama, T., Itoh, K., Frohman, M.A., and Kanaho, Y. (2004) Essential role for phospholipase D2 activation downstream of ERK MAP kinase in NGF-stimulated neurite outgrowth from PC12 cells. *J. Biol. Chem.*, 279:37870-7.

98. Bader, M.-F. and Frohman, M.A. (2005) Phospholipase D: Membrane traffic and Exocytosis. *Experimental Medicine*. 23:39-45.

99. Zhang, Y., Kanaho, Y., Frohman, M.A., and Tsirka, S.E. (2005) Phospholipase D1-promoted release of tissue plasminogen activator facilitates neurite outgrowth. *J. Neuroscience*, 25:1797-805.

100. Huang, P., Altshuller, Y.M., Hou, J.C., Pessin, J.E., and Frohman, M.A. (2005) Insulin-stimulated plasma membrane fusion of the Glut4 glucose transporter is regulated by Phospholipase D1. *Mol. Biol. Cell*, 16:2614–23. PMCID: PMC1142410

101. LaLonde, M.M., Janssens, H., Rosenbaum, E., Choi, S.-Y., Gergen, J.P., Colley, N.J., Stark, W.S., and Frohman, M.A. (2005) Regulation of phototransduction responsiveness and retinal degeneration by a phospholipase D–generated signaling lipid. *J. Cell Biol.* 169:471-9. PMCID: PMC2171926

102. Sasaki, J., Sasaki, T., Yamazaki, M., Matsuoka, K., Taya, C., Shitara, H., Wada, T., Nishio, M., Miyazaki, H., Watanabe, H., Kubo, S., Chiba, T., Murata, S., Frohman, M.A., Tanaka, K., Penninger, J.M., Suzuki, A., Yonekawa, H., and Kanaho, Y. (2005) Regulation of anaphylactic responses by phosphatidylinositol phosphate kinase type I α. *J. Exp. Med.,* 201:859-70. PMCID: PMC2213097

103. Xiao, N., Du, G., and Frohman, M.A. (2005) PrxII functions as a signal-terminator for H_2O_2-activated Phospholipase D1. *FEBS J.*, 272:3929-37.

104. Jenkins, G.M. and Frohman, M.A. (2005) Phospholipase D: A Lipid Centric Review. *Cellular and Molecular Life Sciences*, 62:2305-16.

105. Bostrom P., Rutberg M., Ericsson J., Holmdahl P., Andersson L., Frohman M.A., Boren J., Olofsson S.O. (2005) Cytosolic Lipid Droplets Increase in Size by Microtubule-Dependent Complex Formation. *Arterioscler Thromb Vasc Biol.* 25:1945-51.

106. Singh, A.T.K, Frohman, M.A., and Stern, P.H. (2005) Parathyroid Hormone Stimulates Phospholipase D - Mediated Hydrolysis of Ethanolamine-Containing Phospholipids in UMR-106 Osteoblastic Cells. *Lipids*, 40:1135-40.

107. Corrotte, M., Chasserot-Golaz, S., Huang, P., Du, G., Ktistakis, N., Frohman, M.A., Vitale, N., Bader, M.-F., and Grant, N.J. (2006) Dynamics and function of phospholipase D and phosphatidic acid during phagocytosis. *Traffic,* 7: 365-77.

108. Huang, P., Jiang, Z., Teng S, Wong, Y.C., Frohman, M.A., Chung, S.K., and Chung, S.S.M. (2006) Synergism between phospholipase D2 and sorbitol accumulation in diabetic cataract formation through modulation of Na, K-ATPase activity and osmotic stress. *Invest. Ophthalmol. Visual Sci.* 83:939-48.

109. LaLonde, M., Janssens, H., Yun, S., Crosby, J., Redina, O., Du, G., Olive, V., Choi, S.-Y., Altshuller, Y.M., Gergen, J.P., and Frohman, M.A. (2006) A role for Phospholipase D in Drosophila embryonic cellularization. *BMC Developmental Biology,* 6:60. PMCID: PMC1698916

110. Suzuki, T., Kanai, Y., Hara, T., Sasaki, J., Sasaki, T., Kohara, M., Maehama, T., Taya, C., Shitara, H., Yonekawa, H., Frohman, M.A., Yokoseki, T., and Kanaho, Y. (2006) Crucial role of the small GTPase ARF6 in hepatic cord formation during liver development. *Mol. Cell Biol.* 26:6149-56. PMCID: PMC1592812

111. Andersson L., Bostrom P., Rutberg M., Ericsson J., Marchesan, D., Magnusson, B., Ruiz, M., Asp, L., Huang, P., Frohman M.A., Boren J., Olofsson S.O. (2006) PLD1 and ERK2 regulate cytosolic lipid droplet formation. *J. Cell Sci.* 119:2246-5.

112. Cai, D., Netzer, W.J., Zhong, M., Lin, Y., Du, G., Frohman, M.A., Foster, D.A., Sisodia, S.S., Gorelick, F.S., Xu, H. and Greengard, P. (2006) Presenilin-1 (PS1) utilizes Phospholipase D1as a negative regulator of □-amyloid formation. *Proc Natl Acad Sci U.S.A.* 103:1941-6. PMCID: PMC1413665

113. Vicogne, J., Vollenweider, D., Smith, J., Huang, P., Frohman, M. and Pessin, J. (2006) Asymmetric phospholipid distribution drives *in vitro* reconstituted SNARE-dependent membrane fusion. *Proc Natl Acad Sci U.S.A.*, 103:14761-6. PMCID: PMC1595425

114. Choi, S.-Y., Huang, P., Jenkins, G.M., Chan, D.C., Schiller, J., and Frohman, M.A. (2006) A common signaling lipid requirement for Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. *Nature Cell Biology*, 8:1255-62.

115. Du. G., Yonekubo, J., Zeng, Y., Osisami, M., and Frohman, M.A. (2006) Design of small-hairpin RNA expression vectors based on miRNAs and RNA splicing. *FEBS J.,* 273:5421-7.

116. Scotto-Lavino, E., Du, G., and Frohman, M.A. 5' end cDNA amplification using classic RACE (2006) *Nature Protocols*, 1:2555-62.

117. Scotto-Lavino, E., Du, G., and Frohman, M.A. Using 3' RACE to clone cDNA ends. (2006) *Nature Protocols*, 1:2742-5.

118. Scotto-Lavino, E., Du, G., and Frohman, M.A. Amplification of 5' end cDNA with 'New RACE'. (2006) *Nature Protocols*, 1:3056-61.

119. Choi, S.-Y., Gonzalvez, F., Jenkins, G.M., Slomianny, C., Chretien, D., Arnoult, D., Petit, P.X., and Frohman, M.A. (2007) Cardiolipin deficiency releases cytochrome *c* from the inner mitochondrial membrane and accelerates stimuli-elicited apoptosis. *Cell Death and Differentiation*, 14:597-606.

120. Kobayashi-Nakano, A., Yamazaki, M. Unoki, T., Hongu, T., Murata, C., Taguchi, R., Katada, T., Frohman, M.A., Yokozeki, T., and Kanaho, Y. (2007) Role of activation of PIP5Kγ661 by AP-2 complex in synaptic vesicle endocytosis. *EMBO J.* 26:1105-16. PMCID: PMC1852847

121. Zhao, C., Du, G., Frohman, M.A., and Bar-Sagi, D. (2007) Phospholipase D2-generated PA couples EGFR stimulation to Ras activation by Sos. *Nature Cell Biology*, 9:707-712.

122. Kanaho, Y., Nakayama, K., Frohman, M.A., and Yokozeki, T. (2007) Regulation of Phosphatidylinositol 4-Phosphate 5-Kinase activity by partner proteins. *Methods in Enzymology*. 434:155-69.

123. Huang, P. and Frohman, M.A. (2007) The potential for phospholipase D as a new therapeutic target. *Expert Opinion on Therapeutic Target* 11:707-16.

124. Pannequin, J., Delaunay, N., Darido, C., Maurice, T., Crespy, P., Frohman, M.A., Matter, K., Bourgaux, J.-F., Bali, J.P., and Hollande, F. (2007) Phosphatidylethanol accumulation promotes intestinal hyperplasia by inducing ZONAB-mediated cell density increase in response to chronic ethanol exposure. *Molecular Cancer Research*, 5:1147-57.

125. Zeniou-Meyer, M., Zabari, N., Ashery, U., Chasserot-Golaz, S., Haeberlé, A.-M., Demais, V., Bailly, Y., Gottfied, I., Nakanishi, H., Neiman, A.M., Du, G., Frohman, M.A., Bader, M.-B., and Vitale, N. (2007) PLD1 production of phosphatidic acid at the plasma membrane promotes exocytosis of large dense-core granule at a late stage. *J. Biol. Chem.*, 282:21746-57.

126. Yang, J.-S., Gad, H., Lee, S., Mironov, A., Zhang, L., Beznoussenko, G.V., Valente, C., Turacchio, G., Bonsra, A.N., Du, G., Baldanzi, G., Graziani, A., Bourgoin, S., Frohman, M.A., Luini, A., and Hsu, V.W. (2008) COPI vesicle fission: a role for phosphatidic acid and insight into Golgi maintenance. *Nature Cell Biology,* 10:1146-53. PMCID: PMC2756218

127. Su, W. and Frohman, M.A. (2009) Methods for Measuring the Activity and Expression of Phospholipase D. In: *Lipid-mediated Signal Transduction* (Rosenberger, T. and Murphy, E.J., Editors), CRC Press Methods in Signal Transduction.

128. Hauff, K.D., Choi, S.-Y., Frohman, M.A. and Hatch, G.M. (2009) Cardiolipin synthesis is required to support human cholesterol biosynthesis from palmitate upon serum removal in Hela cells. *The Canadian Journal of Physiology and Pharmacology*, 87: 813-820.

129. Su, W., Chen, Q., and Frohman, M.A. (2009) Targeting phospholipase D with small-molecule inhibitors as a potential therapeutic approach for cancer metastasis. *Future Oncology*, 5:1477-1486. PMCID: PMC2814819

130. Du, G. and Frohman, M.A. (2009) A lipid-signaled myosin phosphatase surge disperses cortical contractile force early in cell spreading. *Molecular Biology of the Cell,* 20:200-8. PMCID: PMC2613090

131. Yeku., O., Scotto-Lavino, E., and Frohman, M.A. (2009) Identification of Alternative transcripts using Rapid Amplification of cDNA Ends (RACE). *Methods in Molecular Biology,* 590:279-94.

132. Su, W. and Frohman, M.A. Phospholipase D. (2009) In Ralph A. Bradshaw and Edward A. Dennis, editors: Handbook of Cell Signaling 2nd edition, Oxford: Academic Press, 144:1167-1176.

133. Nishikimi, A., Fukuhara,H., Su, W., Hongu, T., Takasuga, S., Mihara, H., Cao, Q., Sanematsu, F., Kanai, M., Hasegawa, H., Tanaka, Y., Shibasaki, M., Kanaho, Y., Sasaki, T., Frohman, M.A., and Fukui, Y. (2009) Sequential Regulation of DOCK2 Dynamics by Two Phospholipids during Neutrophil Chemotaxis. *Science*, 324:384-7.

134. Frohman, M.A. Phospholipase D2. (2009) Targeted Proteins Database, Current Biodata. In press.

135. Osisami, M., Huang, H., and Frohman, M.A. (2009) Monitoring insulin-stimulated production of signaling lipids at the plasma membrane. *Methods in Molecular Biology*. 590:61-70.

136. Huang, H. and Frohman, M.A. (2009) Lipid signaling on the mitochondrial surface. *Biochim Biophys Acta.* special Phospholipase D issue. Eds. Frohman M.A., and Cockcroft, S. 1791:839-44. PMCID: PMC2749073

137. Cockcroft S, Frohman M.A. (2009) Special issue on Phospholipase D. *Biochim Biophys Acta.* 1791:837-8.

138. Su, W., Yeku, O., Olepu, S., Genna, A., Park, J.-S., Ren, H., Du, G., Gelb, M.H., Morris, A.J., and Frohman, M.A. (2009) FIPI, a pan-Phospholipase D pharmacological inhibitor that alters cell spreading and inhibits chemotaxis. *Mol. Pharm.* 75: 437-46. PMCID: PMC2684902

139. Lu, J., Fan, T., Zhao, Q., Zeng, W., Zaslavsky, E., Chen, J.J., Frohman, M.A., Golightly, M.G., Madajewicz, S., and Chen, W.-T. (2010) Isolation of circulating epithelial and tumor progenitor cells with an invasive phenotype from breast cancer patients. *International Journal of Cancer*, 126:669-683. PMCID: PMC2795034

140. Scotto-Lavino, E., Garcia-Diaz, M., Du, G., and Frohman, M.A. (2010) The basis for the isoform-specific interaction of Myosin Phosphatase subunits Protein Phosphatase 1c b and myosin phosphatase targeting subunit 1. *J. Biol. Chem.*, 285:6419-24. PMCID: PMC2825437

141. Schauss, A.C., Huang, H., Choi, S.-Y., Xu, L., Bilodeau, P., Zunino, R., Rippstein, P., Frohman, M.A., and McBride, H.M. (2010) A cell-free biomolecular complementation assay for mitochondrial fusion. *Biomed Central Biology,* 8:100. PMCID: PMC2919466.

142. Nishioka, T., Frohman, M.A., Matsuda, M., and Kiyokawa, E. (2010) Heterogeneity of phosphatidic acid levels and distribution at the plasma membrane in living cells as visualized by a forster resonance energy transfer (fret) biosensor. *J. Biol. Chem.*, 285:35979-87. PMCID: PMC2975220.

143. Tsukahara, T., Tsukahara, R., Cheng, Y.-H., Zhang, C., Balazs, L., Du, G., Frohman, M.A., Baker, D., Parrill, A.L., Uchiyama, A., Kobayashi, T., Murakami-Murofushi, K. and Tigyi, G. (2010) The novel second messenger Cyclic Phosphatidic Acid Negatively Regulates the Nuclear Hormone Receptor PPARγ. *Molecular Cell,* 39:421-32.

144. Elvers, M., Stegner, D., Hagedorn, I., Kleinschnitz, C., Braun, A., Kuijpers, M.E.J., Boesl, M., Chen, Q., Heemskerk, J.W.M., Stoll, G., Frohman, M.A., and Nieswandt, B. (2010) Impaired integrin aIIbb3 activation and shear-dependent thrombus formation in mice lacking phospholipase D1. *Science Signaling*, 3:1-10.

145. Frohman, M.A. and Tsirka, S.E. (2010) Polymerase Chain reaction: specialized applications. In: Nature Encylopedia of Life Sciences (Robertson, S., Editor). Macmillan, London. In press.

146. Yanase, Y., Carvou, N., Frohman, M.A., and Cockcroft, S. (2010) Reversible bleb formation in mast cells stimulated with antigen is Ca^{2+}-calmodulin–dependent and bleb size is regulated by ARF6. *Biochem J.*, 425:179-193.

147. Yin, H., Gui, Y., Du, G., Frohman, M.A., and Zheng, X.-L. (2010) Dependence of phospholipase D1 multi-monoubiquitination on its enzymatic activity and palmitoylation. *J. Biol. Chem*, 285:13580-8. PMCID: PMC2859519

148. Qin, H., Frohman, M.A. and Bollag, W.B. (2010) Phospholipase D2 Mediates Acute Aldosterone Secretion in Response to Angiotensin II in Adrenal Glomerulosa Cells. *Endocrinology*, 151:2162-70. PMCID: PMC2869249

149. Huang, H., Choi, S.-Y., and Frohman, M.A. (2010) A Quantitative Assay for Mitochondrial Fusion using *Renilla* Luciferase Complementation. *Mitochondrion*, 10:559-566. PMCID: PMC2922018.

150. Dall'armi C, Hurtado-Lorenzo A, Tian H, Morel E, Nezu A, Chan RB, Yu WH, Robinson KS, Yeku O, Small SA, Duff K, Frohman MA, Wenk MR, Yamamoto A, Di Paolo G. (2010) The Phospholipase D1 Pathway Modulates Macroautophagy. *Nature Communications*, 1:142.

151. Ren, H., Federico, L., Huang, H., Sunkara, M., Drennan, T., Frohman, M.A., Smyth, S.S. and Morris. A.J. (2010) A phosphatidic acid-binding/nuclear localization motif determines lipin 1 function in metabolism and adipogenesis. *Mol. Biol. of the Cell*, 21:3171-81. *PMCID:PMC2938383*

152. Frohman, M.A. (2010) Mitochondria as integrators of signal transduction and energy production in cardiac physiology and disease. *J. Mol. Med.*, 88(10):967-970. PMCID: PMC2964372

153. Frondorf, K., Henkels, K.M., Frohman, M.A., and Gomez-Cambronero, J. (2010) Phosphatidic acid is a leukocyte chemoattractant through S6K activation. *J. Biol. Chem.*, 285:15837-47. PMCID: PMC2871451

154. Yoon, M.-S., Du, G., Backer, J.M., Frohman, M.A. and Chen J. (2011) hVps34 activates PLD1 in an amino acid-sensing mTORC1 pathway. *J. Cell Biology*, 195:435-47.

155. Yeku., O. and Frohman, M.A. (2011) Rapid Amplification of cDNA Ends (RACE). *Methods in Molecular Biology,* 703:107-22.

156. Huang, H., Gao, Q. Peng, X.X., Choi, S.-Y., Sarma, K., Ren, H., Morris, A.J., and Frohman, M.A. (2011) piRNA-associated germline nuage formation and spermatogenesis require MitoPLD pro-fusogenic mitochondrial-surface lipid signaling. *Dev. Cell*, 20:376-387. PMCID: PMC3061402.

157. Huang, P., Yeku, O., Zong, H., Tsang, P., Su, W., Xu, X., Teng, S., Osisami, M., Kanaho, Y., Pessin, J.E., and Frohman, M.A. (2011) Phosphatidylinositol-4-Phosphate-5-Kinase □ Deficiency Alters Dynamics of Glucose-Stimulated Insulin Release to Improve Glucohomeostasis and Decrease Obesity in Mice. *Diabetes*, 60:454-463.

158. Su, W. and Frohman, M.A. Phospholipase D. (2011) In Ralph A. Bradshaw and Edward A. Dennis, editors: Transduction Mechanisms in Cell Signaling 2nd edition, Oxford: Academic Press.

159. Peng, X.X. and Frohman, M.A. (2012) *Mammalian phospholipase D physiological and pathological roles*. *Acta Physiologica*, 204:219-26.

160. Huang, H. and Frohman, M.A. (2012) Visualizing mitochondrial lipids and fusion events in mammalian cells. Methods in Cell Biology. 108:131-45. PMCID: PMC3722243.

161. Jäntti, M.H., Putula, J., Somerharju, P., Frohman, M.A., Kukkonen, J.P. (2012) OX_1 orexin/hypocretin receptor activation of phospholipase D. *Biochemical Journal*, 165:1109-23.

162. Gao, Q. and Frohman, M.A. Roles for the lipid-signaling enzyme MitoPLD in mitochondrial dynamics, piRNA biogenesis, and spermatogenesis. (2012) *Biochemistry and Molecular Biology Reports*, 45:7-13. PMCID: PMC3760789.

163. Frohman, M.A. (2012) New insights into the role of lipid signaling in spermatogenesis. *INTERNATIONAL REVIEW OF CELL AND MOLECULAR BIOLOGY.*

164. Osisami, M., Ali, W. and Frohman, M.A. A role for Phospholipase D3 in myotube formation. (2012). *PLoS One*, 7(3): e33341.

165. Chen, Q., Hongu, T., Sato, T., Zhang, Y., Ali, W., Cavallo, J.-A., van der Velden, A., Tian, H., Di Paolo, G., Nieswandt, B., Kanaho, Y., and Frohman, M.A. (2012) Key roles for the lipid signaling enzyme phospholipase D1 in the tumor microenvironment during tumor angiogenesis and metastasis. *Science Signaling*, 5:ra79.
    - with accompanying Podcast in *Science Signaling*; highlighted in *Nature Cancer Reviews* (Jan. 2013)

and in the *Cancer Discovery Research Watch* by the **A**merican **A**ssociation for **C**ancer **R**esearch (Nov. 15th, 2012).

166. Yang, C.-Y. and Frohman, M.A. (2012) MITOCHONDRIA: Signaling with Phosphatidic Acid. *The International Journal of Biochemistry & Cell Biology*. 44:1346-1350. PMCID: PMC3380155.

167. Ali, W. and Frohman, M.A. (2013) Phospholipase D. In *Encyclopedia of Biological Chemistry, 2nd edition.* (eds. Lennarz, W. and Lane, M.D.), Academic Press, Waltham, MA, 3:472-477.

168. Shuo Li, Marie-Paule Lefranc, Vincent Corbin, Douglas Freeman, Véronique Giudicelli, Eltaf Alamyar, Jean-Pierre Scheerlinck, Magdalena Plebanski, Paul Cameron, Michael Frohman, Bruce Loveland, Andrew Lucas, Stephen Turner, Scott R. Burrows, Anthony T Papenfuss, John J. Miles and Eric J Gowans. (2013) High throughput sequencing analysis of natural regulatory and conventional T cell receptor repertoires during human H1N1 challenge. *Nature communications,* 4:2333.

169. Ali, W., Chen, Q., Delgiorna, K.E., Su, W., Hall, J.C., Hongu, T., Tian, H., Kanaho, Y., Di Paolo, G., Crawford, H.C., and Frohman, M.A. (2013) Deficiencies of the lipid-signaling enzymes phospholipase D1 and D2 alter cytoskeletal organization, macrophage phagocytosis, and cytokine-stimulated neutrophil recruitment. *PLoS One*, 8.e55325.

170. Altshuller, A., Gao, G., and Frohman, M.A. (2013) A C-terminal transmembrane anchor targets the nuage-localized spermatogenic protein Gasz to the mitochondrial surface. *ICRN Cell Biology*, pii: 707930.

171. Stegner, D., Thielmann, I., Kraft, P. Frohman, M.A., Stoll, G., and Nieswandt, B. (2013) Pharmacological inhibition of phospholipase D protects mice from occlusive thrombus formation and ischemic stroke. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 33:2212-7.

172. Baba, T., Kashiwagi, Y., Arimitsu, N., Kogure, T., Edo, A., Maruyama, T., Nakao, K., Hamamoto, H., Sekimizu, K., Nakanishi, H., Kinoshita, M., Frohman, M.A., Yamamoto, A., and Tani, K. (2014) Phosphatidic acid (PA)-preferring phospholipase A1 regulates mitochondrial dynamics. *J. Biol. Chem.*, 289:11497-511.

173. Ha, E.E. and Frohman, M.A. (2014) Regulation of Mitochondrial Morphology by Lipids. *Biofactors*, 40(4):419-24.

174. Teng S., Stegner D., Chen Q., Hongu T., Hasegawa H., Chen L., Kanaho Y., Nieswandt B., Frohman M.A., Huang P. (2014) Phospholipase D1 facilitates second phase myoblast fusion and skeletal muscle regeneration. ***Molecular Biology of the Cell***, E14-03-0802.

175. Frohman, M.A. (2014) Polymerase Chain reaction: specialized applications. In: Nature Encyclopedia of Life Sciences (Robertson, S., Editor). Macmillan, London.

176. Zhang, Y. and Frohman, M.A. (2014) Cellular and physiological roles for Phospholipase D1 in cancer. *J. Biol. Chem.*, 289(33):22567-74. PMID: 24990946.

177. Tsai, Y.Y., Rainey, W., Pan, Z.-Q., Frohman, M.A., Choudhary, V and Bollag, W.B. (2014) Phospholipase D Activity Underlies Very Low Density Lipoprotein (VLDL)-induced Aldosterone Production in Adrenal Glomerulosa Cells. *Endocrinology*, 155(9):3550-60.

178. Akiyama, M., Hasegawa, H., Hongu, T., Frohman, M.A., Harada, A., Sakagami, H., and Kanaho, Y. (2014) Trans-regulation of oligodendrocyte myelination by neurons through small GTPase Arf6-regulated secretion of fibroblast growth factor-2. *Nature Communications*, 5:4744.

179. Choudhary, V., Olala, L.O., Qin, H., Helwa, I., Pan, Z.-Q., Tsai, Y.-Y., Frohman, M.A., Kaddour-Djebbar, I., and Bollag, W.B. (2015) Aquaporin-3 re-expression induces differentiation in a phospholipase D2-dependent manner in aquaporin-3 knockout mouse keratinocytes. *Journal of Investigative Dermatology*, 135:499-507.

180. Mallipattu SK, Horne SJ, D'Agati V, Narla G, Liu R, Frohman MA, Dickman K, Chen EY, Ma'ayan A, Bialkowska AB, Ghaleb AM, Nandan MO, Jain MK, Daehn I, Chuang PY, Yang VW, He JC. (2015) Krüppel-like factor 6 regulates mitochondrial function in the kidney. *J. Clinical Investigation*, 125:1347-61. PMCID:PMC4362257

181. Frohman, M.A. (2015) Role of Mitochondrial lipids in guiding fission and fusion. Journal of Molecular Medicine, 93:263-9. PMCID:PMC4334719

182. Frohman, M.A. (2015) The Phospholipase D superfamily as therapeutic targets. *Trends in Pharmacological Sciences*, 36(3):137-144. PMCID:PMC4355084

183. Nelson, R. and Frohman, M.A. (2015) Phospholipases: Central Role in Lipid Signaling and Disease. *Journal of Lipid Research*, 56(12):2229-37. PMCID:PMC4655994

184. Birey, F., M. Kloc, M. Chavali, I. Hussein, M. Wilson, D.J. Christoffel, T. Chen, M.A. Frohman, J.K. Robinson, S.J. Russo, A. Maffei, and A. Aguirre (2015) Genetic and stress-induced loss of NG2+ glia triggers emergence of depressive-like behaviors through reduced secretion of FGF2. *Neuron*, 88(5):941-56.

185. Adachi, Y., Itoh, K., Cerveny, K.L., Suzuki, T., Macdonald, P., Frohman, M.A., Ramachandran, R., Iijima, M., and Sesaki, H. (2016) Coincident Phosphatidic Acid Interaction Restrains Drp1 in Mitochondrial Division. *Molecular Cell,* 63(6):1034-43.

186. Lee, H.-J., Li, F.-C., Ruan, D., Powers, R.S., Frohman, M.A., and Chang, C.-H. (2016) The DNA damage transducer RNF8 facilitates cancer chemoresistance and progression through Twist activation. *Molecular Cell,* 63(6):1021-33.

187. Philip, F., Ha E.E., Seeliger, M.A. and Frohman, M.A. (2017) Measuring Phospholipase D Enzymatic Activity Through Biochemical and Imaging Methods. In *Enzymology at the Membrane Interface, Methods in Enzymology*. 583:309-325.

188. Frohman, M.A. (2016) Phospholipase D. In *Encyclopedia of Biological Chemistry, 3rd edition.* (eds. Lennarz, W. and Lane, M.D.), Academic Press, Waltham, MA.

189. Nelson, R.K., Gadbery, J., Abedeen, D., Sampson, N. and Frohman, M.A. (2017) Phospholipase D2 loss increases blood pressure via negative regulatory effects on the endothelial nitric oxide synthase pathway. *Scientific reports. 7, 9112 doi:10.1038/s41598-017-09852-4.*

190. Roth E. and Frohman MA. (2017) Phospholipase D in cancer pathways. Advances in Biological Regulation, pii: S2212-4926(17)30178-1.

191. Ganesan R, Henkels KM, Wrenshall LE, Kanaho Y, Paolo GD, Frohman MA, Gomez-Cambronero J. (2018) Oxidized LDL phagocytosis during foam cell formation in atherosclerotic plaques relies on a PLD2-CD36 functional interdependence. J Leukoc Biol. 103(5):867-883. PMID: 29656494

192. Li, H., Guo, C, Kim-Holzapfel, D., Li, W., Altshuller, Y., Schroeder, B., Liu, W., Meng, Y., French, J., Takemaru, K.-I., Frohman, M.A., Jia, S. (2018) Fast, volumetric live-cell imaging using high-resolution light-field microscopy. ***Biomedical Optics Express*** 10(1):29-49. PMCID: PMC6363205.
- first winner of the *Biomedical Optics Express* Best Paper Prize

193. Frohman, M.A. (2018) Phospholipase D. Reference Module in Biomedical Sciences. Elsevier. 17-Dec-18 doi: 10.1016/B978-0-12-801238-3.11144-4.

194. Salazar, C. and Frohman, M.A. (2019) Prospects for PLD Inhibition in Cancer and Thrombotic Disease Handb Exp Pharmacol. doi: 10.1007/164_2019_244. PMID: 31541323.

195. Barisano, D., and Frohman, M.A. (2020) Roles for Phospholipase D1 in the Tumor Microenvironment. (2020) Adv. Exp. Med. Biol., 1259:77-87.

196. Frohman, M.A. (2020) Polymerase Chain reaction: specialized applications. In: Nature Encyclopedia of Life Sciences (Robertson, S., Editor). Macmillan, London.

197. Bowling, F.Z., Salazar, C.M., Bell, J.A., Huq, T.S., Frohman, M.A., and Airola, M.V. (2020) Crystal structure of human PLD1 provides insight into activation by PI(4,5)P_2 and RhoA. ***Nature Chemical Biology***, 16(4):400-407.

198. Lahrouchi, N, Postma, AV, Salazar, CM,... Frohman, MA*, and Bezzina, BR* (2021) Bi-allelic loss-of-function variants in *PLD1* cause congenital right-sided cardiac valve defects and neonatal cardiomyopathy. ***J. Clinical Investigation***, 131(5):e142148. *, co-corresponding authors

199. Bowling F.Z., Frohman M.A., and Airola M.V. (2021) Structure and regulation of human phospholipase D. Adv Biol Regul., 79:100783. PMID: 33495125

200. Frohman, Michael A. (2021) Phospholipase D. In: Jez Joseph (eds.) Encyclopedia of Biological Chemistry, 3rd Edition. vol. 2, pp. 766–775. Oxford: Elsevier.

201. Solana, A. and Frohman, M. (2022) Phospholipase D. Journal of Lipid Research, in press.

202. Li, F., Wu, Z., Gao, Y.q., Bowling, F.Z., Franklin. J.M., Hu, C., Suhandynata, R.T., Frohman, M.A., Airola, M.V., Zhou, H., Guan, K.L. (2022) Regulation of PLD1 and PI4KB by ARF GTPases revealed by proximal interaction network. *EMBO Journal,* 41(17):e110698. PMID: 35844135.

203. Cavallo, J.-A., Trocchia, C., Fleischman, S., Ha, E.E., Obeid, L.M., Shroyer, K.R., Snider, A.J., Frohman, M.A. Phospholipase D1 ablation suppresses neoangiogenesis and pulmonary metastasis in MMTV-PyMT-driven breast cancer. *under review*

204. Hala A. Addassi, Irena Krga, Fernando Villarreal, Joseph F. LaComb, Michael A. Frohman, Karen Matsukuma, and Gerardo G. Mackenzie. Inhibition of phospholipase D1 reduces pancreatic carcinogenesis in mice through a FAK-dependent mechanism. (2024) *In press.*

204. Salazar, C.M., Roth, E.E., Rosati, B. McKinnon, D., and Frohman, M.A. Loss of Phospholipase D1 and D2 decreases cancer progression in the MMTV-PyMT breast cancer model through changes in the tumor epithelium revealed by single cell RNA sequencing. *In preparation.*

Book Reviews: Choi, S-Y and Frohman, M.A. (2002) Gene Cloning and DNA Analysis: An Introduction. 4th edition, by T. A. Brown; Blackwell Science Inc. In **The Quarterly Review of Biology.**

PATENTS: Patent Number 6,043,073 "Phospholipase D Polypeptide and DNA Sequences"; March 28th, 2000. Michael A. Frohman, JoAnne Engebrecht, Andrew J. Morris

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/260106507

FDA Guidance on Pyrogens and Endotoxin

Article *in* GMP Review · July 2013

CITATIONS
2

READS
6,661

1 author:



Tim Sandle
The University of Manchester
766 PUBLICATIONS **1,661** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Laboratory Management View project



Endotoxin View project

All content following this page was uploaded by Tim Sandle on 10 February 2014.

The user has requested enhancement of the downloaded file.

FDA guidance on pyrogens and endotoxin testing

by Tim Sandle

The Limulus Amebocyte Lysate (LAL) assay is the compendial test for the examination of bacterial endotoxin in pharmaceutical products, as well as intermediates, medical devices, and in water. Although there are various emerging alternative methods, there are three described in the pharmacopoeia: gel-clot, chromogenic and turbidometric (the latter two are photometric methods). Aside from pharmacopoeia texts, only the US Food and Drug Administration (FDA) has elected to have a guidance document for the LAL test. This document was first issued in 1987, revised in 1991, and withdrawn in 2011[1].

In 2012 the FDA issued their long awaited guidance relating to Pyrogen and Endotoxin testing[2]. Unlike the previous guidance, which went into considerable descriptive detail on test me, the new FDA guidance takes the form of 'questions and answers'.The document is titled: "Guidance for Industry: Pyrogen and Endotoxins Testing: Questions and Answers" and it is dated June 2012.The introduction to the FDA document reads:

"This guidance provides recommendations for biological product, drug, and device firms on FDA's current thinking concerning the testing recommendations and acceptance criteria in the United States Pharmacopeia (USP) Chapter on Bacterial Endotoxins Test, USP Chapter."

The important phrase from the introduction is the connection between 'FDA current thinking' and the United States Pharmacopeia chapter for conducting endotoxin testing using the Limulus Amebocyte Lysate (LAL) assay chapter <85>[3].

The reason for the FDA guidance

To understand the reason for the FDA guidance, the history of the harmonisation process needs to be briefly considered. In 2001, the United States Pharmacopeia chapter on the endotoxin test was harmonised with the equivalent European Pharmacopoeia monograph for the endotoxin test (2.6.14), along with the relevant section in the Japanese Pharmacopoeia (JP 4.01). This process was overseen by ICH Q4B Expert Working 107 Group (EWG) ([4]). The two most important decisions from the harmonisation were:

- Any of the three techniques can be used for the test. In the event of doubt or dispute, the gel-clot limit test should be used to make the final decision on compliance for the product being tested.
- The Endotoxin Reference Standard should be calibrated to the current WHO (World Health Organization) International Standard for Endotoxin.

A further revision to the harmonised chapters took place in 2005 and again in 2010, with the publication of USP edition 33 and EP edition 7.0 (January 2011)[5]. Even though the chapters are harmonised the text is not identical.

As part of this harmonisation process most of the troubleshooting parts of the respective monographs were removed because those involved with the harmonisation process could not reach agreement. This led to the European Pharmacopoeia developing its own, non-mandatory guidance monograph (5.1.10 "Guidelines for using the Test for Bacterial Endotoxins")[6]. Similarly, and at a slower pace, the FDA undertook to issue its own revised guidance.

The LAL test is subject to assay variation and the samples under test can cause assay interference to the degree that assay can be difficult to validation; furthermore, there are complications in relation to the setting of enodotoxin limits and for setting the maximum valid dilution (a means to ensure that a sample is not overly diluted to the extent that any endotoxin cannot be measured against the endotoxin limit). These various issues have lead to the need for additional guidance in relation to the assay and, at present, international agreement cannot be reached so that this guidance can be incorporated into a globally accepted pharmacopoeial monograph. Hence, the FDA's need for its own guidance document.

The 2012 FDA Guidance

For those manufacturing products with the USA or for distribution with the USA, familiarity with the FDA pyrogens and endotoxin guidance is important. Unlike the supporting 5.1.10 chapter in the European Pharmacopoeia, the FDA chapter does not seek to cover the entire subject of pyrogen and endotoxins testing. The aim of the guidance is to address "those issues that may be subject to misinterpretation and are not covered in compendial procedures or in currently available guidance documents".

As indicated above, the guidance takes the form of a series of questions and answers.

The first substantive question addresses the issue of sampling plans for in-process and finished product testing. This is to ensure that the samples taken for testing are representative of the product (which is a GMP requirement of the Code of Federal Regulations: CFR 211.160). Here the FDA direct manufacturers to consider the risk from endotoxins and the potential for contamination across the pharmaceutical manufacturing process, from raw materials, through intermediate products, to finished products. The important points to consider are where are endotoxin removal steps in place? (and does this require verification through testing?);

Seikagaku Corporation Exhibit 2004 - Page 43 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

and where to prolonged process hold times occur and, as part of quality control, is there a requirement to conduct a LAL test? (this relates to a risk from Gram-negative bacteria growing in sufficient numbers).

The second important point relates to re-testing. This to relates to a situation when the LAL test records a suspect out-of-specification (OOS) result. The FDA guidance allows one re-test to be conducted at a greater dilution (provided that this additional dilution does not exceed the maximum valid dilution). If the re-test confirms the OOS then the sample or lot should be rejected. The guidance additionally notes that all test criteria and test limits must be established prior to the test being conducted.

The third part considers sample storage and sample handling given that the detection of endotoxins is affected by the conditions that any sample is subjected to. The section also addresses the importance of sample mixing to ensure that any endotoxin which may have adhered to the surface of a container is recovered. Many of these issues should be addressed as part of method development.

The fourth part addresses an issue that has been in and out of favour for several decades: whether it is permissible to pool samples and to run a composite test. The FDA guidance comes down in favour of sample pooling. This would, for example, that several vials of a final product could be mixed together and one test conducted. There are dangers with this approach, not least the concern that if a positive result was detected understand where in the batch this occurred would not be possible. There is also a danger of endotoxin being masked (and the impact upon the maximum valid dilution). A further concern is that the more manipulations that are undertaken, the greater the chance there is of cross-contamination occurring. It would be preferable, contrary to the FDA guidance, to test individual samples.

The fifth part of the document, as a point of encouragement for those interested in rapid microbiological methods, considered alternative methods to the LAL assay. Here the FDA states that there are no obstacles to alternative methods provided that they are validated and that the recommendations of the USP general validation chapter <1225> are observed. If there is a dispute in relation to any alternative method, the LAL gel-clot method remains the final arbiter.

Examples of alternative endotoxin tests include the use of recombinant lysate (lysate not taken from the horseshoe crab). Recombinant lysates, based on factor-C, are offered by several LAL manufacturers. A second example is the Monocyte Activation Test (MAT), for which there is a European Pharmacopoeia monograph (chapter 2.6.30). the MAT is based on the in vitro activation of human monocytoid cells by pyrogens, which leads to the release of proinflammatory cytokines that are determined by ELISA[7].

The FDA guidance provides an outline of how a pharmaceutical manufacturer could make the transition from the pharmacopoeial methods to an alternative method. This involves running several samples challenged with bacterial endotoxin and noting the level of recovery and level of detection. There is also an interesting note about seasonal variability and assessing whether this has any impact upon levels of endotoxin obtained throughout the manufacturing process.

The sixth key section of the guidance document explains why the FDA no longer offer recommended endotoxin limits. The FDA explains that do to so would not be scientifically valid and instead directs the pharmaceutical manufacturer towards a calculation for endotoxin limits in the USP. This calculation takes into account the average weight of a patient and the quantity of the drug that they are likely to be infused with over a one hour period.

The seventh main section embraces the FDA's current interest with the principles of Quality by Design. This relates back to the earlier section on the sampling plan, and considers how the risk from endotoxins should be considered for each new product stream or process.

The eighth section considers the rabbit pyrogen test, a test already redundant throughout most of Europe. The FDA position is that if a USP monograph requires the pyrogen test to be performed then it is permissible to substitute the endotoxin test as an alternative method. A note of caution is added in relation to proving, robustly, that the endotoxin method is suitable in relation to test interference.

The next two sections address issues relating veterinary products and medical devices. With medical devices the endotoxin limit stated is 0.5 EU/mL (based on extraction volume), with separate limits applied to products that contact cerebrospinal fluid. A note of caution is added about the complexities of validating the extraction method, with requires the medical device to rinsed with endotoxin free water

The eleventh section provides some guidance in relation to the maximum valid dilution and how a product which requires dilution in order to overcome test interference is diluted as little as possible. For example, if the maximum valid dilution of a product is set at 1: 100, and test interference is found to occur at any dilution up to 1: 10, but not at 1: 20, then the appropriate test dilution to use is 1: 20 rather than testing at 1: 100.

The twelfth and final section addresses endotoxin standards, where the importance of suing a control endotoxin traceable to the international standard is emphasised.

Seikagaku Corporation Exhibit 2004 - Page 44 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Summary

Although the FDA guidance is less comprehensive than the supporting chapter in the European Pharmacopoeia it provides an important indication of the Agency's thinking on the subject and, unlike the EP guidance, it includes topical subjects like Quality by Design. By adopting a Q&A format the FDA guidance is clear and straightforward to read and the questions posed generally cover most of those considered by laboratory staff tasked with carrying out he test and QA personnel who have a responsibility in releasing products subject to endotoxin testing.

References

1 FDA (1987). Guideline on validation of the limulus amebocytelysate test as an end-product endotoxin test for human and animal parenteral drugs, biological products, and medical devices. Washington DC, U.S. Food and Drug Administration. www.fda.gov/downloads/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidance/Blood/UCM080966.pdf

2 FDA (2012). Guidance for Industry Pyrogen and Endotoxins Testing: Questions and Answers, U.S. Department of Health and Human Services Food and Drug Administration: Bethesda, MD, USA. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM310098.pdf

3 United States Pharmacopeia (2010) Bacterial Endotoxins Test. Chapter <85>. USP 33: pp. R65-R69. United States Pharmacopeial Convention, Rockville, MD

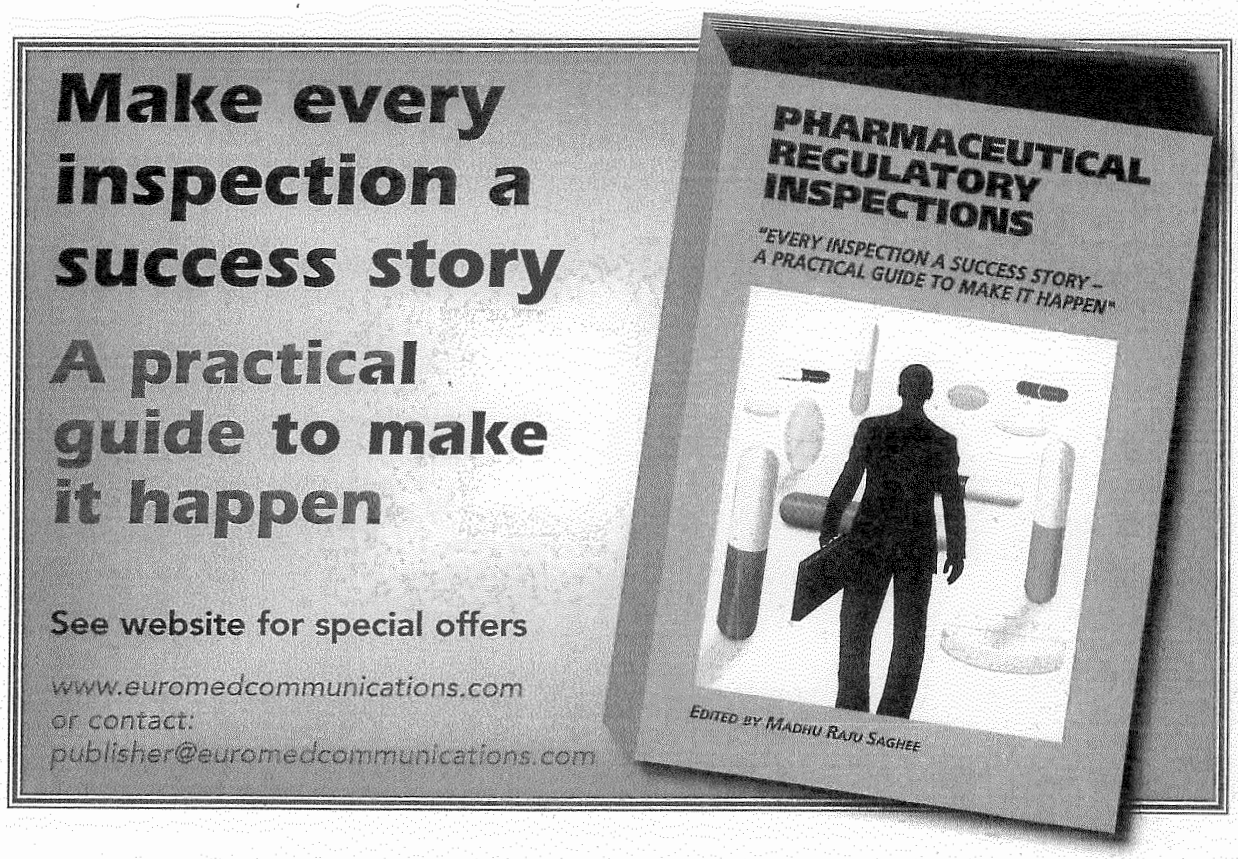
4 CDER (2010). Q4B Evaluation and Recommendation of Pharmacopoeial Texts for Use in the ICH Regions, Annex 14: Bacterial Endotoxins Test General Chapter, Center for Drug Evaluation and Research, Food and Drug Administration: Bethesda, MD, USA.

5 European Pharmacopoeia 2.6.14. Bacterial Endotoxins (2010) European Pharmacopoeia 7.0, p.171-175. European Pharmacopoeia Commission, Strasbourg, France

6 European Pharmacopoeia 5.1.10. Guidelines for using the Test for Bacterial Endotoxins (2010), European Pharmacopoeia 7.0, p.520-523. European Pharmacopoeia Commission, Strasbourg, France

7 Brandenburg K, Howe J, Gutsman T, and Garidel P. The expression of endotoxic activity in the Limulus test as compared to cytokine production in immune cells. *Curr Med Chem* 2009;**16**:2653-60.

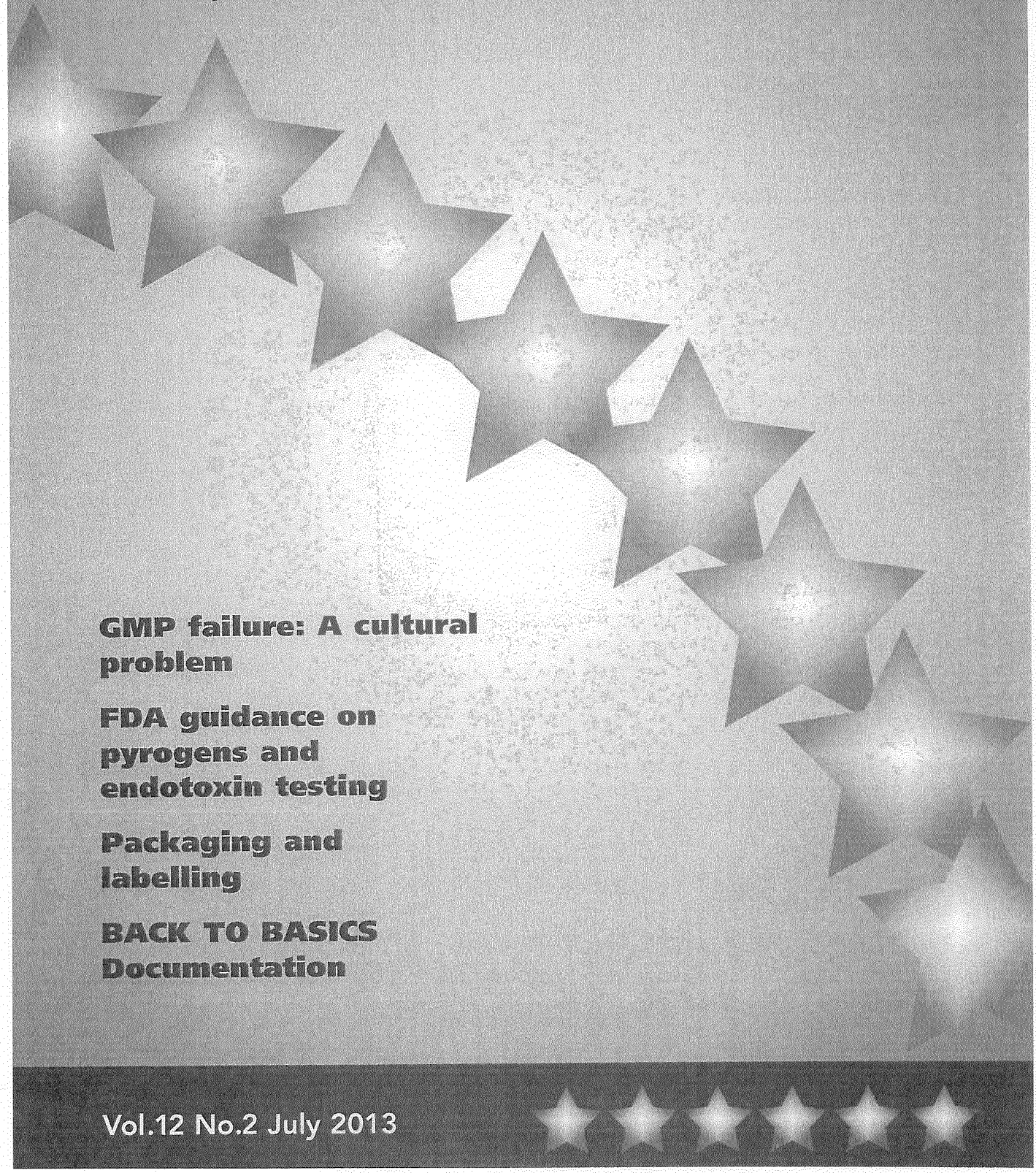
Tim Sandle is the Head of Microbiology at Bio Products Laboratory, UK
Email: tim.sandle@bpl.co.uk
Website:www.pharmamicro.com



Seikagaku Corporation Exhibit 2004 - Page 45 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

gmp review

analysing international pharmaceutical regulations



GMP failure: A cultural problem

FDA guidance on pyrogens and endotoxin testing

Packaging and labelling

BACK TO BASICS Documentation

Vol.12 No.2 July 2013

View publication stats

european PHARMACEUTICAL manufacturer

Advantages of recombinant Factor C based endotoxin testing

17 NOVEMBER 2015 12:44

Lakiya Wimbish, Lonza, discusses the benefits of recombinant Factor C based endotoxin testing



Endotoxins are lipopolysaccharides found in the outer membrane of Gram-negative bacteria. These pyrogenic components can elicit an immune response if administered via the blood stream through contaminated parenteral drugs or other medical devices. Hence,

Seikagaku Corporation Exhibit 2004 - Page 47 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

bacterial endotoxin testing (BET) needs be carried out for all such products in order to prevent any adverse effects.

In the 1950s, Fred Bang discovered that horseshoe crab (*Limulus polyphemus, Tachypleus gigas* and *T. tridentatus*) blood reacts in the presence of endotoxin. This is due to the presence of Factor C, a component that activates the clotting cascade (Figure 1) and forms the basis of the Limulus Amebocyte Lysate (LAL) and Tachypleus Amebocyte Lysate (TAL) assays used in BET today.

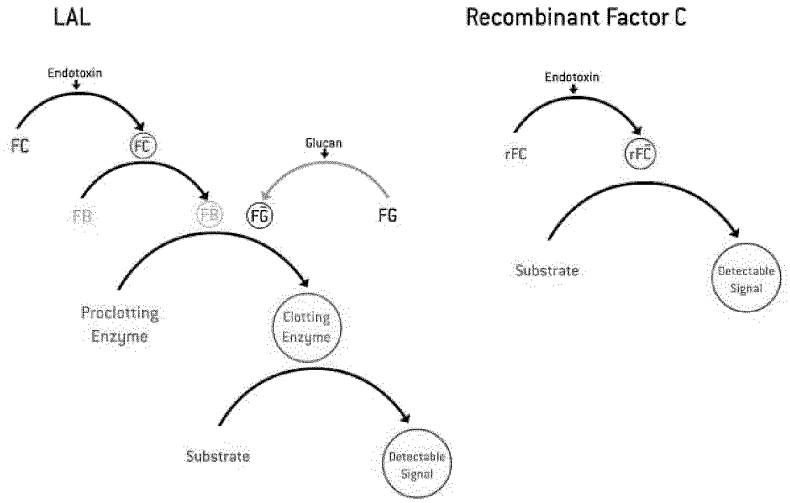


Figure 1: Comparing LAL turbidimetric and chromogenic assays to fluorescence-based Pyrogene recombinant Factor C assay.

The industry is heavily reliant upon the supply of LAL/TAL for endotoxin testing. However, scientists have now developed a recombinant version of Factor C (rFC) that is comparable to the natural equivalent. One of the reasons the industry has perhaps been reluctant to adopt this alternative is the fact that the pharmacopoeial regulations currently necessitate an additional validation procedure to be carried out. However, this is not as complicated or time-consuming as it may appear.

The need for alternative tests

The need for the widespread adoption of an alternative assay is becoming increasingly important due the fact that the demand for LAL/TAL may soon exceed the supply. This is due to an increasing

Seikagaku Corporation Exhibit 2004 - Page 48 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

number of new biologics being developed, the growing impact of personalised medicine and the maturing Asia-Pacific markets. In addition, as with any natural resource, crab populations are at risk of mismanagement and loss. Therefore, adopting rFC-based methods would not only meet European and global initiatives to reduce animal use in testing, but would also decrease the demand made on the crabs, a finite resource.

In addition to negating the need for an animal source, rFC products, such as Lonza's fluorescence-based Pyrogene rFC assay, perform comparably or better than LAL/TAL-based methods for some applications, and offer a number of additional benefits to the user. These include improved lot-to-lot consistency, enhanced endotoxin specificity, statistically more robust spike recovery, ease-of-use and a suitable sensitivity range (0.005 EU/ml – 5 EU/ml).

Validating the alternative

So what is holding back manufacturers from adopting rFC-based assays? According to the United States Pharmacopeia (USP) rFC-based assays are listed as 'alternative methods', and must be validated as per USP <1225> or International Conference on Harmonisation (ICH) Q2B. This means that their use requires extended validation over and above the compendial methods, to prove the alternative tests meet the intended analytical application. While these steps do require some additional effort, these can be accomplished in one to three days of following a well-structured protocol, alleviating any concerns of significantly lengthening the testing workflow.

The procedure includes four steps. The initial qualification involves an assessment of the reagents and testing equipment to establish whether the current system is functioning properly. The second step determines the level of dilution required to overcome product inhibition or enhancement. Thirdly, the alternative method must be validated and should show that the rFC is capable of achieving comparable results to the current method. In order for the regulators to evaluate the validity of the proposed method, submissions should

Seikagaku Corporation Exhibit 2004 - Page 49 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

contain the following: 1) a rationale for using the test, such as the fact that it reduces a reliance on animals and offers enhanced endotoxin specificity and reproducibility; 2) a complete description of the proposed analytical procedure; and 3) data elements that demonstrate the analytical performance of the procedure. The latter should include the following parameters: accuracy, precision, specificity, linearity, range, detection limit, quantitation limit, and robustness. The fourth and final step of the validation process requires a product-specific validation to ensure it is possible to generate consistent results using the chosen method.

Conclusion

Clearly, there is a need for LAL/TAL assay alternatives due to the growing demand on resources. Fortunately, rFC-based methods are available that offer a number of benefits to the user. Although defined as an alternative by the pharmacopoeia, these tests are simply a fluorescence-based version of the existing LAL/TAL assay and do not require much additional effort to implement. Hence, companies should not be dissuaded from adopting these methods. To make the switch even easier, companies such as Lonza have compiled pre-developed documentation to assist with regulatory submissions and can also help users to incorporate the technology in their testing workflow. It seems likely that rFC will become the go-to solution for BET in the future, and many forward-thinking manufacturers are already adopting it into their processes.

17 NOVEMBER 2015 12:44

CONNECT WITH US
Contact Us
Media Kit

FOLLOW US

RAPID NEWS GROUP PUBLICATIONS
Medical Plastics News
Med-Tech

CORPORATE
Terms & Conditions
Privacy Policy
Part of Rapid News

Seikagaku Corporation Exhibit 2004 - Page 50 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023



Innovation News

TCT Magazine

Manufacturing Quality

Group

© Rapid Life Sciences Ltd, a Rapid News Group Company. All Rights Reserved.

BUILT WITH METRO PUBLISHER™

Yeku, O., & Frohman, M. A. (2010). Rapid Amplification of cDNA Ends (RACE). Methods in Molecular Biology, 107–122. doi:10.1007/978-1-59745-248-9_8

Chapter 8

Rapid Amplification of cDNA Ends (RACE)

Oladapo Yeku and Michael A. Frohman

Abstract

Rapid Amplification of cDNA ends (RACE) provides an inexpensive and powerful tool to quickly obtain full-length cDNA when the sequence is only partially known. Starting with an mRNA mixture, gene-specific primers generated from the known regions of the gene and non-specific anchors, full-length sequences can be identified in as little as 3 days. RACE can also be used to identify alternative transcripts of a gene when the partial or complete sequence of only one transcript is known. In the following sections, we outline details for rapid amplification of 5′ and 3′ cDNA ends using the "new RACE" technique.

Key words: RACE, new RACE, alternative transcripts.

1. Introduction

The advent of microarrays and powerful bioinformatics analysis has not only led to the discovery of new genes but has also provided tools for analysis of existing genes (1–3). Full or partial sequences of newly discovered genes can be aligned against entire organism genomes in search of homology or any other clue that can yield insight into the identity or function of the gene. There are instances, however, where bioinformatics data is unavailable or incomplete. In these instances, rapid amplification of cDNA ends (RACE) can be used to identify full-length sequences of a gene if only part of its sequence is known (4–6). RACE can be used to amplify both the 5′ and the 3′ ends of genes yielding valuable information such as the location of transcription initiation sites, *cis*-acting elements, and the localization and stability of the transcript. By using 3′ and 5′ RACE, full-length cDNA sequences can be cloned from partial sequences in as little as 3 days.

H. Nielsen (ed.), *RNA*, Methods in Molecular Biology 703,
DOI 10.1007/978-1-59745-248-9_8, © Springer Science+Business Media, LLC 2011

Seikagaku Corporation Exhibit 2004 - Page 52 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Since the initial description of RACE (7), many labs and commercial companies have adapted and modified the protocol to increase specificity and user friendliness (8–18) (*see* **Note 1**). Amplification of the 5′ end can generally be divided into two classifications; classic RACE and "new RACE." The same principles underlie all RACE protocols and they differ only in terms of specificity, ease of use and cost. While the principles for classic RACE will be touched on, it has been thoroughly described elsewhere for both 5′ (19) and 3′ (20) end amplification. This protocol will address the identification of alternative transcripts using the "new RACE" protocol (21).

RACE primarily utilizes RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction) and PCR to amplify the ends of transcripts starting with mRNA and cDNA, respectively (**Fig. 8.1**). To perform RACE, the partial or a complete sequence of the mRNA of interest has to be known. This is required to generate the Gene Specific Primers (GSPs) that will be used for the amplifications. A second set of primers that extend from the unknown end of the of the message back to the known region of the 3′ end is provided by the poly (A) tail (or an appended homopolymer), while an appended homopolymer tail is used for the 5′ end. Herein lies the first and perhaps most important difference between "classic" and "new" RACE (22, 23). In classic RACE (19), the homopolymer is appended after the mRNA is reverse transcribed, whereas in "new" RACE, the homopolymer is appended before the reverse transcriptase reaction (*see* **Fig. 8.2**). This simple difference eliminates the amplification of non-full length products, which greatly improves the ability to identify transcription start sites. After the mRNA is reverse transcribed to cDNA, the 5′ and 3′ ends are amplified using two nested PCR reactions using gene-specific primers and primers derived from the sequence of the RNA oligonucleotide (appended homopolymer). The use of sequential nested PCR reactions is important to reduce the amplification of unwanted products, since in each reaction, only one of the primers is specific for the gene of interest and the other binds to all cDNAs present in the starting mixture (by comparison, standard PCR reactions employ two gene-specific primers).

The starting mRNA mixture is dephosphorylated with shrimp alkaline phosphatase (SAP). This dephosphorylates degraded and uncapped (non full-length) mRNAs, leaving the full length mRNAs with methylated "G" caps intact (24). The methylated "G" cap is then removed with tobacco acid pyrophosphatase (TAP). Treatment with TAP exposes the phosphorylated 5′ end of the mRNA and prepares it for ligation to the linker or homopolymer. A short synthetic RNA, prepared by in vitro transcription of a linearized plasmid (25), is ligated to the uncapped 5′ and 3′ end using T4 RNA ligase (*see* **Fig. 8.3**). Degraded or other

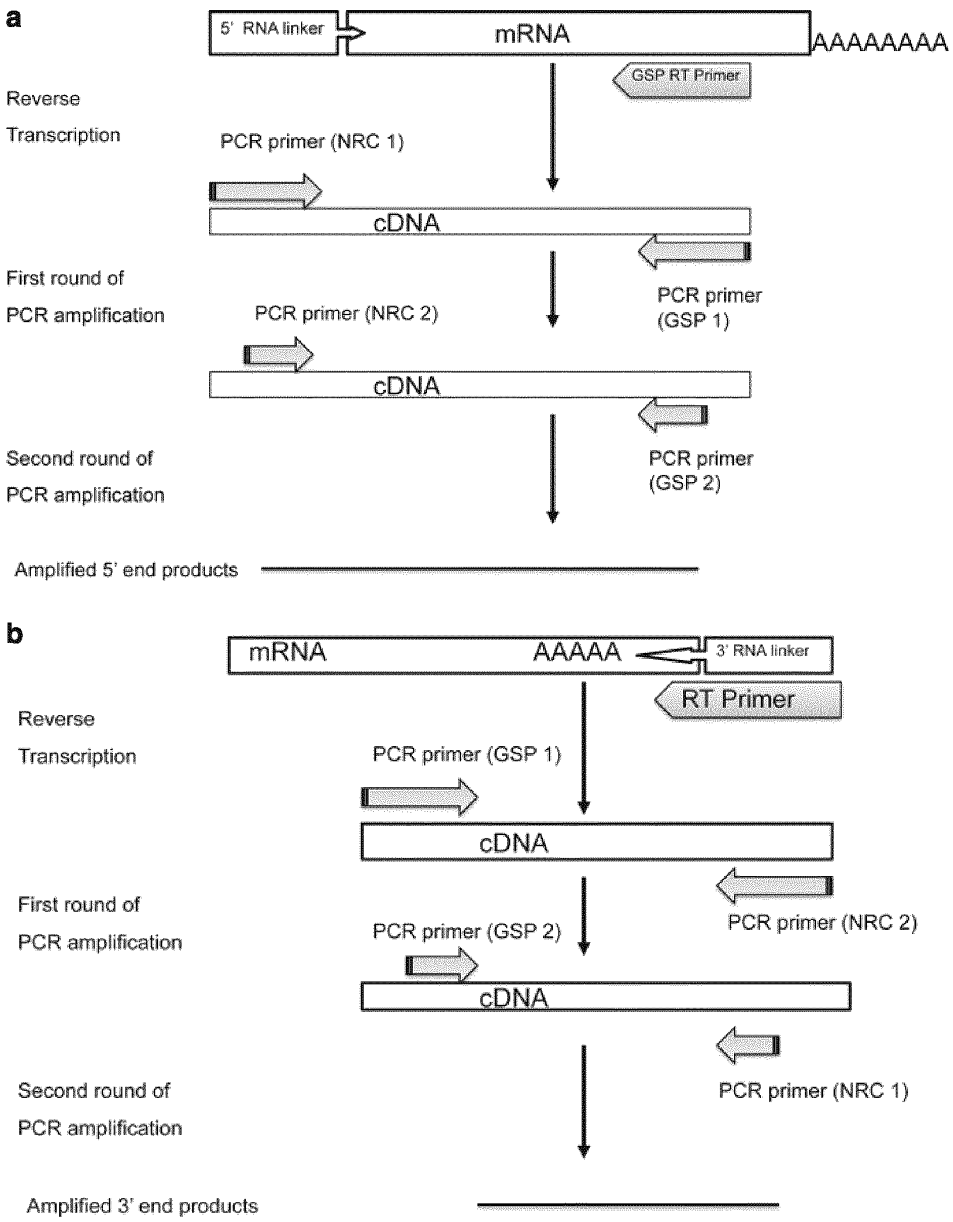


Fig. 8.1. Figure showing the general schematic for Rapid Amplification of cDNA Ends (RACE) starting with an mRNA pool. **a** Amplification of the 5′ end. See text for full details. **b** Amplification of the 3′ end. Note that the NRC 1, NRC 2 and NRC 3 primers are the reverse compliments of sequences described in text. RT: Reverse Transcriptase. GSP: Gene Specific Primer. NRC: New RACE Complement.

non-full length mRNA will not be ligated with the synthetic oligonucleotide, since they are dephosphorylated. The mRNA-RNA oligonucleotide hybrids are then reverse-transcribed using a GSP. Full-length cDNA generated from the previous step then contains specific sequences at the 5′ and 3′ ends that can be used to amplify the full-length cDNA ends. Two nested PCR reactions are used to amplify the full-length cDNA product with high specificity (*see* **Fig. 8.3**).

Seikagaku Corporation Exhibit 2004 - Page 54 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

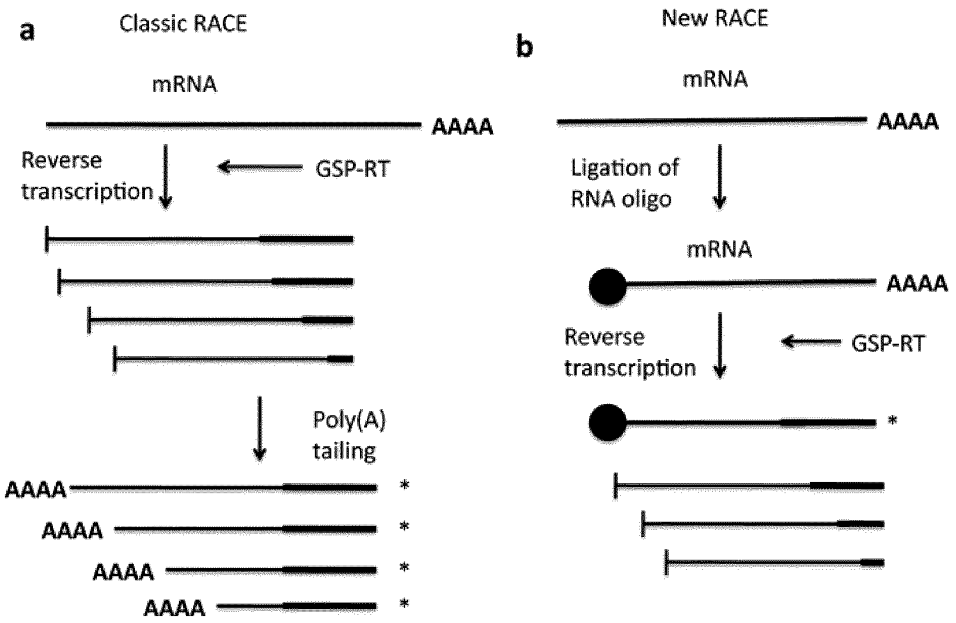


Fig. 8.2. Comparison between "classic RACE" and "new RACE." **a** In "classic RACE," the homopolymer (PolyA tail) is added after the reverse transcription process. This anchor provides specificity for the amplifications downstream of the GSP-reverse transcription step. **b** In "new RACE," the homopolymer is appended before the reverse transcription takes place. This ensures full length products from the onset. Figure reproduced from citation (21). **http://www.nature.com/nprot/journal/v1/n6/images/nprot.2006.480-F3.jpg**.

2. Materials

2.1. Dephosphorylation of Degraded RNA

1. RNA sample (TAP-treated and untreated). All reagents must be RNase free.
2. Phosphatase buffer (10×): 0.1 M Tris-HCl (pH 7.5 at 37°C), 0.1 M $MgCl_2$, and 1 mg/mL BSA.
3. DTT (0.1 M).
4. RNasin (40 U/μL).
5. Shrimp Alkaline Phosphatase SAP (1 U/μL; Fermentas).

2.2. Decapping of Intact RNA

1. Tobacco Acid Pyrophosphatase (TAP) buffer (10×): 0.5 M sodium acetate (pH 6.0), 10 mM EDTA, 1% β-mercaptoethanol, and 0.1% Triton X-100
2. TAP (5 U/μl; Epicentre)
3. TE buffer: 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA, pH 8.0
4. RNA spin column (Qiagen)
5. DTT (0.1 M)
6. RNasin (40 U/μL).

2.3. Preparation of RNA Oligonucleotide

1. TE buffer: 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA, pH 8.0
2. Plasmid template DNA for transcribing RNA oligonucleotide (*see* **Note 2**)
3. Restriction enzymes and buffers
4. Proteinase K (5 mg/mL stock solution (100×))
5. H_2O treated with diethylpyrocarbonate (DEPC)
6. Transcription buffer (5×): Provided with the enzyme *or* 200 mM Tris-HCl (pH 8.0), 40 mM $MgCl_2$, 10 mM spermidine, 250 mM NaCl
7. rUTP solution (10 mM)
8. rATP solution (10 mM)
9. rCTP solution (10 mM)
10. rGTP solution (10 mM)
11. DNA-dependent RNA polymerase (20 U/μL)
12. RNase-free DNase I (0.5 Kunitz units)
13. DTT (0.1 M)
14. RNasin (40 U/μL).

2.4. RNA Oligonucleotide–Cellular RNA Ligation

1. RNA sample (TAP-treated and untreated)
2. Ligation buffer (10×): 500 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP (pH 7.8 at 25°C) (*see* **Note 3**)
3. RNA oligonucleotide (*see* **Note 4**)
4. TP (2 mM)
5. T4 RNA ligase (New England BioLabs) (20 U/μL).

2.5. Reverse Transcription

1. SuperScript II reverse transcriptase (Invitrogen) (200 U/μL). Reverse-transcription buffer (5×) is supplied with the transcriptase.
2. dNTP solution (containing all four dNTPs, each at 10 mM)
3. Gene-specific antisense primer for 5′ end RACE or reverse complement of NRC3 (new Race Compliment primer) for 3′ end RACE (20 ng/μL)
4. RNase H (2 U/μL)
5. DTT (0.1 M)
6. RNasin (40 U/μL)
7. TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.

2.6. PCR Amplifications

1. Hercules Hot-Start polymerase buffer (10×; Stratagene). Do not add additional nucleotides if they are already contained in the buffer.

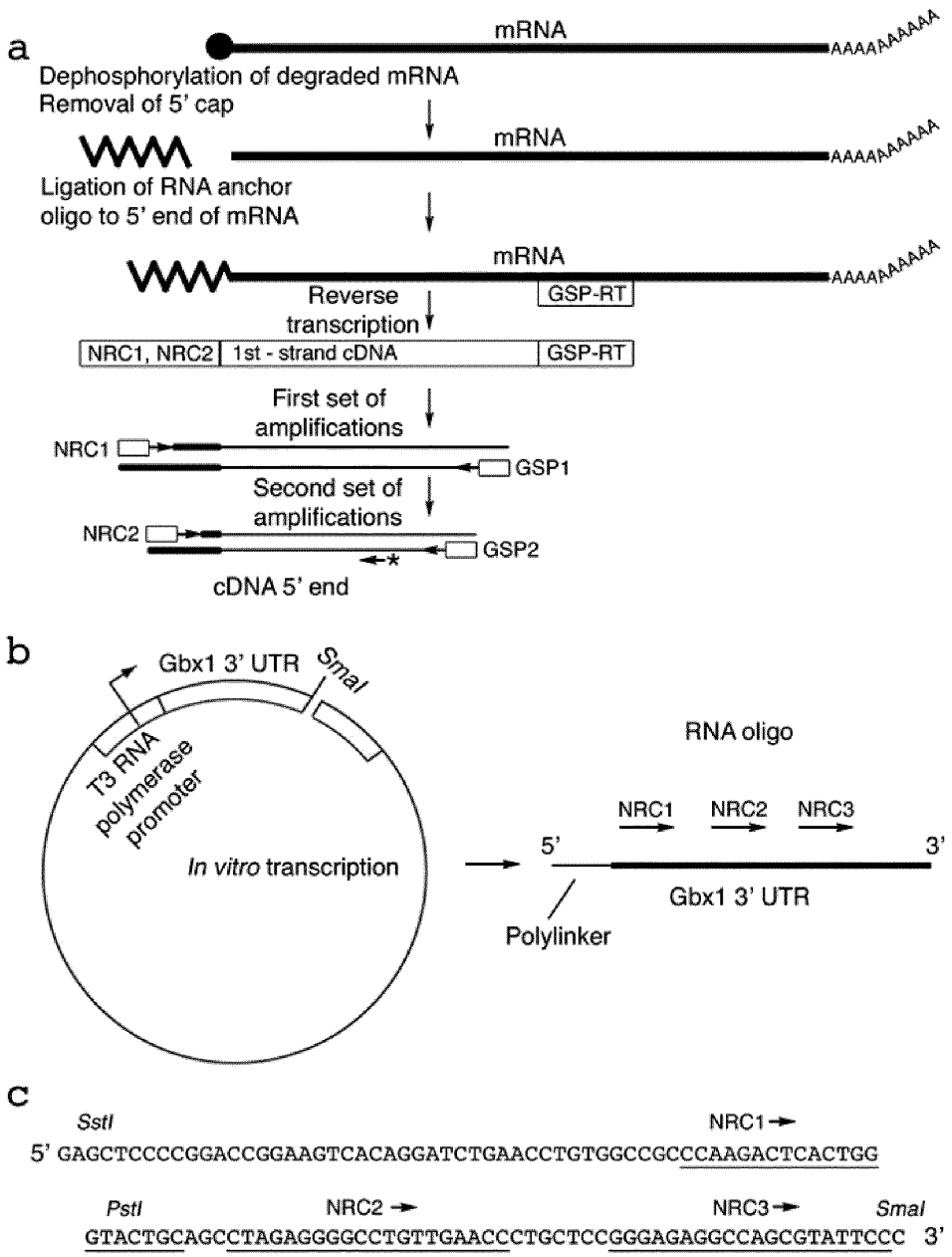


Fig. 8.3. New RACE overview. **a** Outline of steps involved in the amplification of cDNA 5′ ends. See text for detailed explanation. **b** Plasmid map and primers used for the example in the text. **c** In vitro transcription of the T3 plasmid in (**b**) produces a 132-nt product. The sequences for NRC1, NRC2 and NRC3 used for the generation of corresponding primers are underlined. Figure reproduced from citation (21). **http://www.nature.com/nprot/journal/v1/n6/fig_tab/nprot.2006.479_F1.html**.

2. Hercules Hot-Start polymerase (Stratagene). Hot-start protocol is strongly recommended (*see* **Note 5**).
3. User-defined gene-specific oligonucleotide primers GSP1, GSP2, NRC1, and NRC2 to clone the 5′ end and the reverse compliment of GSP1, GSP2, NRC1, NRC2, and NRC3 to

Seikagaku Corporation Exhibit 2004 - Page 57 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

clone the 3′ end (*see* **Note 6** for primer design considerations and **Fig. 8.3** for details of primers NRC1 and NRC2).

2.7. DNA/RNA Purification and Electrophoresis

1. Agarose gel (1%) in TAE buffer
2. Ethidium bromide staining bath (0.5 μg/mL ethidium bromide in electrophoresis buffer; prepared from a stock solution of 10 mg/mL in water)
3. Sodium acetate: 3 M, pH 5.2
4. Phenol/chloroform (1:1 (vol/vol))
5. Chloroform
6. Ethanol
7. Microcon spin filters (Millipore).

3. Methods

If large amounts of RNA are used, it is strongly recommended that the quality control steps be performed. Small samples of the RNA can be run on an agarose gel to check for degradation. Aliquots can then be stored at –80°C for future experiments. The quantities presented in this protocol are the ideal amounts of RNA to be used. Reactions can be scaled down in the event that RNA quantities are limited.

3.1. Dephosphorylation of Degraded RNA

1. Combine reagents listed in **Table 8.1** in a sterile microfuge tube. Follow the manufacturers guidelines regarding the use of SAP.
2. The reaction should be incubated for 1 h at 37°C. Dephosphorylation of uncapped mRNA prevents degraded fragments from participating in the subsequent ligation step.

Table 8.1
Dephosphorylation of degraded RNA

Component	Amount	Final
RNA	50 μg	50 μg
10X phosphatase buffer	5 μL	1×
DTT (0.1 M)	0.5 μL	1 mM
RNasin (40 U/μL)	1.25 μL	50 U
SAP (1 U/μL)	3.5 μL	3.5 U
H_2O	to 50 μL	–

Seikagaku Corporation Exhibit 2004 - Page 58 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

3. Incubate the reaction for 15 min at 65°C to inactivate SAP, then spin briefly in a microcentrifuge. At this point, the products can be stored at –80°C.
4. To visually confirm that the RNA remained intact during the dephosphorylation step, 2 μg (2 μL) of RNA should be analyzed by electrophoresis through a 1% agarose gel (TAE buffer) adjacent to a lane containing 2 μg of the original RNA preparation. Upon ethidium bromide staining, degraded RNA will be present as a low molecular weight smear on the gel rather than a discrete band of high molecular weight. If evidence of RNA degradation is observed, then fresh reagents should be prepared and the step repeated.

3.2. Decapping of Intact RNA

1. Combine and mix the reagents listed in **Table 8.2** in a sterile microfuge tube. Note that the quantity of TAP proposed is sufficient for the reaction.
2. Incubate the reaction for 1 h at 37°C; then add 200 μL TE buffer to stop the reaction.
3. Spin columns provided in any RNA extraction kit or from varied companies can be used for separation of the RNA from the reaction components. Resuspend the RNA in 40 μL H_2O. It is permissible to store the products at this point at –80°C.
4. 2 μg of RNA should be analyzed by electrophoresis (through a 1% agarose gel in TAE buffer) adjacent to a lane containing 2 μg of the original RNA preparation; stain the gel with ethidium bromide and visually confirm that the RNA remained intact during the decapping step. A low molecular weight smear on the gel indicates the presence of RNA fragments, whereas a discrete band of high molecular weight corresponds to the desired product.

Table 8.2
Decapping of intact RNA

Component	Amount	Final
RNA (obtained in Section 3.1, Step 4)	48 μg	48 μg
TAP buffer (10×)	5 μL	1×
DTT (0.1 M)	0.5 μL	1 mM
RNasin (40 U/μL)	1.25 μL	50 U
TAP (5 U/μL)	1 μL	5 U
H_2O	to 50 μL	–

3.3. Preparation of RNA Oligonucleotide

1. Using digestion by appropriate restriction enzymes and buffers, linearize 25 μg of the plasmid to be transcribed. Ensure that the plasmid is reasonably RNase-free.
2. Eliminate any residual RNase activity in the reaction by treating the digestion for 30 min at 37°C with 50 μg/mL proteinase K. Extract twice with phenol–chloroform and once with chloroform, and collect the DNA by standard ethanol precipitation.
3. Redissolve the template DNA in 25 μL TE buffer, pH 8.0. This yields a final concentration of approximately 1 μg/μL.
4. Mix the transcription reagents in the order listed in **Table 8.3** in a sterile microfuge tube at room temperature (25°C).
5. Incubate for 1 h at 37°C to allow transcription to occur.
6. Remove the DNA template by adding 0.5 Kunitz units of DNase I (RNase-free) for every 20 μL of reaction volume and incubate for 10 min at 37°C.
7. Analyze 5 μL of the test or preparative reaction by electrophoresis through a 1% agarose gel (TAE buffer). Expect to see a diffuse band of about the right size for the expected product (or a bit smaller) in addition to some smearing up and down the gel.
8. Purify the oligonucleotide by extracting the reaction mixture with phenol–chloroform and then chloroform (keeping

Table 8.3
Preparation of RNA oligonucleotide

Component	Amount (test scale)	Final	Amount (preparative scale)	Final
DEPC-treated H_2O	4 μL	–	80 μL	–
Transcription buffer (5×)	2 μL	1×	40 μL	1×
DTT (0.1 M)	1 μL	10 mM	20 μL	10 mM
rUTP (10 mM)	0.5 μL	0.5 mM	10 μL	0.5 mM
rATP (10 mM)	0.5 μL	0.5 mM	10 μL	0.5 mM
rCTP (10 mM)	0.5 μL	0.5 mM	10 μL	0.5 mM
Rgtp (10 mM)	0.5 μL	0.5 mM	10 μL	0.5 mM
Linearized DNA from **Section 3.3 Step 3** (1 μg/μL)	0.5 μL	0.5 μg	10 μL	10.0 μg
RNasin (40 U/μL)	0.25 μL	10 U	5 μL	200 U
RNA polymerase (20 U/μL)	0.25 μL	5 U	5 μL	100 U
Total	10 μL	–	200 μL	–

Seikagaku Corporation Exhibit 2004 - Page 60 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

the aqueous phase each time). Perform buffer exchange and clean-up of the RNA oligonucleotide by diluting it to 1 mL with H_2O and then concentrating it using a Microcon spin filter (pre-rinse the spin filter with H_2O before addition of sample). Rinse twice more with 1 mL of H_2O as above. Note that Microcon 10 spin filters have a cutoff size of 20 nt and Microcon 30 spin filters have a cutoff size of 60 nt. Be sure to use Microcon 10 filters for oligonucleotides smaller than 100 nt, and Microcon 30 spin filters for anything larger.

9. To check for the concentration and integrity of the sample, analyze a second aliquot of the oligonucleotide by electrophoresis through a 1% agarose gel in TAE buffer. The oligonucleotide distribution pattern should look like that in **Step 7** above. A much smaller band likely indicates that degradation has occurred and the procedure should be repeated with fresh material. At this juncture, the product can be stored at –80°C indefinitely.

3.4. RNA Oligonucleotide–Cellular RNA Ligation

1. Set up two sterile microfuge tubes, one with TAP-treated cellular RNA and another with untreated cellular RNA, and add the remaining components for the ligation reaction (*see* **Table 8.4**). The tube with the untreated cellular RNA serves as a negative control.
2. Incubate for 16 h at 17°C. Overnight is permissible.
3. Purify and concentrate the ligation product by spin filtration with a Microcon 100 spin filter (rinsing product three times with RNase-free H_2O; pre-rinse filter with H_2O). The recovered volume after the Microcon filtration should not exceed 20 μL. Products can be stored at –80°C.
4. Analyze one third of the product by electrophoresis through a 1% agarose gel (TAE buffer) to verify the integrity of the

Table 8.4
RNA oligonucleotide–cellular RNA ligation

Component	Amount	Final
Ligation buffer (10×)	3 μL	1×
RNasin (40 U/μL)	0.75 μL	30 U
RNA oligonucleotide[a]	4 μg	4 μg
TAP-treated or untreated RNA	10 μg	10 μg
ATP (2 mM)	1.5 μL	0.1 mM
T4 RNA ligase (20 U/μL)	1.5 μL	30 U
H_2O	to 30 μL	–

[a]RNA oligonucleotides are at a molar excess of 3–6 over target cellular RNA.

Seikagaku Corporation Exhibit 2004 - Page 61 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

ligation. It should look like the previous samples, as discussed (*see* **Section 3.3, Step 7**).

3.5. Reverse Transcription

1. For each experimental and control sample from **Section 3.4, Step 3**, assemble the transcription components listed in **Table 8.5** on ice in a sterile microfuge tube.

Table 8.5
Reverse transcription

Component	Amount	Final
Reverse-transcription buffer (5×)	4 μL	1×
dNTP mixture (containing all four dNTPs, each at 10 mM)	1 μL	471 μM
DTT (0.1 M)	2 μL	9.41 mM
RNasin (40 U/μL)	0.25 μL	10 U
Total	7.25 μL	–

2. In a separate tube, add 20 ng of a gene-specific antisense primer (e.g., GSP-RT to generate the 5′ end and the reverse compliment of NRC3 to generate the 3′ end) to the remaining RNA (about 6.7 μg) from **Section 3.4, Step 3** in 13 μL of H_2O. Incubate for 3 min at 80°C, cool rapidly on ice and centrifuge for 5 s. To run an optional control reaction, set up an identical tube in parallel to which no reverse transcriptase is added in **Step 3**.
3. Add the RNA-primer mix to the reverse-transcription components, subsequently add 1 μL (200 U) of SuperScript II reverse transcriptase, and incubate for 1 h at 42°C and for 10 min at 50°C.
4. Inactivate the reverse transcriptase by incubating at 70°C for 15 min then centrifuge for 5 s.
5. Destroy the RNA template by adding 0.75 μL (1.5 U) of RNase H to the reaction mixture and incubating it for 20 min at 37°C.
6. Dilute the reaction mixture to 100 μL with TE buffer and store at 4°C. Note that the 5′ end oligonucleotide–cDNA pool can be stored indefinitely at 4°C. Avoid storing at –20°C, which could snap some of the cDNA strands with repeated freeze–thawings.

3.6. First-Round Amplification

1. In a sterile 0.2-mL microfuge tube, mix the reagents listed in **Table 8.6** for each experimental and control sample from **Section 3.5, Step 6**.

Table 8.6
First-round amplification (reaction pre-mix)

Component	Amount	Final
Hercules Hot-Start polymerase buffer (10×)	5 μL	1×
dNTP solution (10 mM)	1.0 μL	200 μM
Hercules Hot-Start polymerase	2.5 U	2.5 U
H_2O	to 50 μL	–

2. Add a 1 μL aliquot of the 5′ end oligonucleotide–cDNA pool and 25 pmol each of primers GSP1 and NRC1 (or the reverse compliment of GSP1 and NRC1 if you are cloning the 3′ end). A "no-template" control should also be prepared.
3. Heat in a DNA thermal cycler for 5 min at 98°C to denature the first-strand (and to activate the polymerase). Cool for 2 min to the appropriate annealing temperature (56–68°C), followed by cDNA extension for 40 min at 72°C.
4. Carry out 35 cycles of amplification with a "step" program as listed in **Table 8.7**. Products can be stored at 4°C.

Table 8.7
First-round amplification (heat-block settings)

Cycle number	Denaturation	Annealing	Polymerization-extension
1–30	10 s at 94°C	10 s at 52–60°C	3 min at 72°C
31–35	10 s at 94°C	10 s at 52–60°C	15 min at 72°C; then cool to room temperature

3.7. Second-Round Amplification

1. Dilute a portion of the amplification products from the first round amplification 1:20 in TE buffer. Two rounds of amplification ensure the specificity of the yield by quenching the background amplification. The first round generates a lot of background because only one gene-specific primer is utilized. By using a second gene-specific primer, the specificity of the yield is significantly increased.
2. In a sterile 0.2 mL microfuge tube, the reagents listed in **Table 8.8** should be mixed (on ice).
3. Add a 1-mL aliquot of the diluted first-round amplification products (obtained in **Step 1** above) and 25 pmol each of

Table 8.8
Second-round amplification (reaction pre-mix)

Component	Amount	Final
Hercules Hot-Start polymerase buffer (10×)	5 μL	1×
dNTP solution (10 mM)	1.0 μL	200 μM
Hercules Hot-Start polymerase	2.5 U	2.5 U
H_2O	to 50 μL	–

primers GSP2 and NRC2 (or their reverse compliments for amplifying the 3′ end). Set up a "no-template" control as well.

4. Mix and heat in a DNA thermal cycler for 5 min at 98°C to denature the first-strand products and to activate the polymerase.
5. Carry out 30 cycles of amplification with a "step" program as described in **Table 8.9**. The product can be stored at 4°C.
6. Separate 20% of the products of first- and second-round amplification by electrophoresis through a 1% agarose gel. Specific partial cDNAs can be checked by Southern blot analysis. Information gained from this analysis can be used to optimize the RACE procedure.

Table 8.9
Second-round amplification (heat-block settings)

Cycle number	Denaturation	Annealing	Polymerization-extension
1–29	10 s at 94°C	10 s at 52–68°C	3 min at 72°C
30	10 s at 94°C	10 s at 52–68°C	15 min at 72°C; then cool to room temperature

3.8. Anticipated Results

Depending on the stringency of the amplification cycles, the yield of the desired product could range from less than 1 to 100% of the amplified product. After the second set of amplifications, the expected yield of specific products should be about 100%.

If there are alternative transcripts in the starting pool of mRNA, then several different specific products should be seen after the second round of amplifications. These products are identified as distinct bands when visualized on an ethidium bromide stained gel. Upon sequencing, the presence of TATA or CCAAT,

or other promoter elements, and initiator sites around your gene should identify these products as genuine alternative transcripts.

4. Notes

1. RACE kits are available from Invitrogen, Clontech, and Ambion. These ready-made systems are user friendly and powerful. However they are not suited for every purpose. The investigator might have to optimize the kits using principles outlined in this protocol, e.g., using your own gene-specific primers (GSP) and/or varying the annealing temperature. The entire procedure can be completed in 3 days but will take up to 5 days if the procedure is stopped at the storage points.
2. Adenosine residues are the best "acceptors" for the ligation of the 3′ end of the RNA oligonucleotide to the 5′ end of its target, meaning that the restriction enzyme for linearization should be chosen accordingly (for example, *Nde*I, *Sal*I, or *Xho*I, which terminate in an A at the end of the transcription product). A test transcription should be performed to ensure that all reactions are working properly, and then the reaction can be further "scaled up." The oligonucleotide can be stored indefinitely at –80°C for future experiments. It is also important to synthesize sufficient oligonucleotide to allow for losses due to purification and "spot-checks" of the oligonucleotide during use. Alternatively, the DNA or RNA oligonucleotide can be ordered from a commercial source.
3. 10× T4 RNA ligase buffers supplied by some manufacturers contain too much ATP (25). If the composition contains more than 1 mM ATP, it is strongly recommended that you make your own. You can do this by regenerating the components used in the supplied buffer and adjusting the ATP concentration to 1 mM (final 1× concentration should be 0.1 mM). Alternatively, the 10× buffer described in ref (25) can be used.
4. For the synthesis of the RNA oligonucleotide, it is important to select a plasmid that can be linearized at a site about 100 bp downstream from a T7 or T3 RNA polymerase site. A plasmid containing an insert cloned into the first polylinker site should be used because primers made from palindromic polylinker DNA do not amplify well during PCR. In the scenario presented here, pBS-SK-GBX-1-3′UTR contains the 3′ untranslated region of the mouse Gbx1 gene7 cloned into the SstI restriction site of the

plasmid pBS-SK (Stratagene). This plasmid can be linearized with the restriction enzyme SmaI and transcribed with T3 RNA polymerase to produce a 132-nt RNA oligonucleotide, although as suggested in **Note 2**, using a restriction enzyme that generates a cleavage site that terminates in an A residue would be preferable. Any other readily amplifiable sequence can be substituted, but avoid ones derived from abundant endogenous genes (such as actin) to prevent spurious nonspecific amplification.

5. The presence of nonspecific amplification products could be the result of inappropriate annealing temperature. Using "Touch down PCR" is an effective approach to optimize the annealing temperature. Alternatively, the same optimization can be attained by gradually increasing the annealing temperature in 2°C intervals at each relevant step of the protocol.
6. All the primers used for amplification are derived from the RNA oligonucleotide. It is recommended that primer design is software assisted in order to ensure similar melting temperatures. For instance, Primer3 from the Massachusetts Institute of Technology (**http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi**) can be used to generate suitable candidates.

Acknowledgments

This work was supported by NIH awards GM071520 and GM084251 to MAF and an NIH T32 Medical Scientist Training Grant, an F31 National Research Service Award from the National Institute of Diabetes and Digestive and Kidney Diseases, and the Turner Foundation award to OY.

References

1. Ferreira, E. N., Galante, P. A., Carraro, D. M., de Souza, S. J. (2007) Alternative splicing: a bioinformatics perspective. *Mol Biosyst* 3, 473–477.
2. Kashyap, L., Tabish, M., Ganesh, S., Dubey, D. (2007) Identification and comparative analysis of novel alternatively spliced transcripts of RhoGEF domain encoding gene in C. elegans and C. briggsae. *Bioinformation* 2, 43–49.
3. Seim, I., Collet, C., Herington, A. C., Chopin, L. K. (2007) Revised genomic structure of the human ghrelin gene and identification of novel exons, alternative splice variants and natural antisense transcripts. *BMC Genomics* 8, 298.
4. Allen, R. D., 3rd, Dickerson, S., Speck, S. H. (2006) Identification of spliced gammaherpesvirus 68 LANA and v-cyclin transcripts and analysis of their expression in

vivo during latent infection. *J Virol* 80, 2055–2062.
5. Frohman, M. A. (1994) On beyond classic RACE (rapid amplification of cDNA ends. *PCR Methods Appl* 4, S40–S58.
6. Jarosinski, K. W., Schat, K. A. (2007) Multiple alternative splicing to exons II and III of viral interleukin-8 (vIL-8) in the Marek's disease virus genome: the importance of vIL-8 exon I. *Virus Genes* 34, 9–22.
7. Frohman, M. A., Dush, M. K., Martin, G. R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 85, 8998–9002.
8. Bertling, W. M., Beier, F., Reichenberger, E. (1993) Determination of 5′ ends of specific mRNAs by DNA ligase-dependent amplification. *PCR Methods Appl* 3, 95–99.
9. Borson, N. D., Salo, W. L., Drewes, L. R. (1992) A lock-docking oligo(dT) primer for 5′ and 3′ RACE PCR. *PCR Methods Appl* 2, 144–148.
10. Edwards, J. B., Delort, J., Mallet, J. (1991) Oligodeoxyribonucleotide ligation to single-stranded cDNAs: a new tool for cloning 5′ ends of mRNAs and for constructing cDNA libraries by in vitro amplification. *Nucleic Acids Res* 19, 5227–5232.
11. Fritz, J. D., Greaser, M. L., Wolff, J. A. (1991) A novel 3′ extension technique using random primers in RNA-PCR. *Nucleic Acids Res* 19, 3747.
12. Frohman, M. A. (1989) Creating full-length cDNAs from small fragments of genes: Amlification of rare transcripts using a single gene-specific oligonucleotide primer, in (Innis, M., Gelfand, D., Sninsky, J. and White T. C., Eds.), *PCR Protocols and Applications: A Laboratory Manual.* Academic, New York, NY, pp. 28–38.
13. Frohman, M. A. (1993) Rapid amplification of cDNA for generation of full-length cDNA ends: thermal RACE. *Methods Enzymol* 218, 340–356.
14. Frohman, M. A. M., Martin, G. R. (1989) Rapid amplification of cDNA ends using nested primers. *Tech* 1, 165–173.
15. Jain, R., Gomer, R. H., Murtagh, J. J.,Jr. (1992) Increasing specificity from the PCR-RACE technique. *Biotechniques* 12, 58–59.
16. Monstein, H. J., Thorup, J. U., Folkesson, R., Johnsen, A. H., Rehfeld, J. F. (1993) cDNA deduced procionin. Structure and expression in protochordates resemble that of procholecystokinin in mammals. *FEBS Lett* 331, 60–64.
17. Rashtchian, A., Buchman, G. W., Schuster, D. M., Berninger, M. S. (1992) Uracil DNA glycosylase-mediated cloning of polymerase chain reaction-amplified DNA: application to genomic and cDNA cloning. *Anal Biochem* 206, 91–97.
18. Templeton, N. S., Urcelay, E., Safer, B. (1993) Reducing artifact and increasing the yield of specific DNA target fragments during PCR-RACE or anchor PCR. *Biotechniques* 15, 48–50, 52.
19. Scotto-Lavino, E., Du, G., Frohman, M. A. (2006) 5′ end cDNA amplification using classic RACE. *Nat Protoc* 1, 2555–2562.
20. Scotto-Lavino, E., Du, G., Frohman, M. A. (2006) 3′ end cDNA amplification using classic RACE. *Nat Protoc* 1, 2742–2745.
21. Scotto-Lavino, E., Du, G., Frohman, M. A. (2006) Amplification of 5′ end cDNA with 'new RACE'. *Nat Protoc* 1, 3056–3061.
22. Fromont-Racine, M., Bertrand, E., Pictet, R., Grange, T. (1993) A highly sensitive method for mapping the 5′ termini of mRNAs. *Nucleic Acids Res* 21, 1683–1684.
23. Liu, X., Gorovsky, M. A. (1993) Mapping the 5′ and 3′ ends of Tetrahymena thermophila mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic Acids Res* 21, 4954–4960.
24. Volloch, V., Schweitzer, B., Zhang, X., Rits, S. (1991) Identification of negative-strand complements to cytochrome oxidase subunit III RNA in Trypanosoma brucei. *Proc Natl Acad Sci USA* 88, 10671–10675.
25. Tessier, D. C., Brousseau, R., Vernet, T. (1986) Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. *Anal Biochem* 158, 171–178.

Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 8998–9002, December 1988
Biochemistry

Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer

(polymerase chain reaction/5′ and 3′ cDNA ends/cDNA cloning/low-abundance mRNAs/*int-2* gene)

MICHAEL A. FROHMAN, MICHAEL K. DUSH, AND GAIL R. MARTIN

Department of Anatomy, University of California, San Francisco, CA 94143

Communicated by Peter H. Duesberg, September 9, 1988

ABSTRACT **We have devised a simple and efficient cDNA cloning strategy that overcomes many of the difficulties encountered in obtaining full-length cDNA clones of low-abundance mRNAs. In essence, cDNAs are generated by using the DNA polymerase chain reaction technique to amplify copies of the region between a single point in the transcript and the 3′ or 5′ end. The minimum information required for this amplification is a single short stretch of sequence within the mRNA to be cloned. Since the cDNAs can be produced in one day, examined by Southern blotting the next, and readily cloned, large numbers of full-length cDNA clones of rare transcripts can be rapidly produced. Moreover, separation of amplified cDNAs by gel electrophoresis allows precise selection by size prior to cloning and thus facilitates the isolation of cDNAs representing variant mRNAs, such as those produced by alternative splicing or by the use of alternative promoters. The efficacy of this method was demonstrated by isolating cDNA clones of mRNA from *int-2*, a mouse gene that expresses four different transcripts at low abundance, the longest of which is ≈2.9 kilobases. After <0.05% of the cDNAs produced had been screened, 29 independent *int-2* clones were isolated. Sequence analysis demonstrated that the 3′ and 5′ ends of all four *int-2* mRNAs were accurately represented by these clones.**

Despite the development of numerous cDNA cloning strategies (1–5), obtaining full-length cDNA copies of low-abundance mRNAs remains a formidable task. We describe here a simple, rapid, and efficient cDNA cloning strategy that is based on the DNA polymerase chain reaction (PCR) technique developed by Saiki *et al.* (6). PCR employs two oligonucleotide primers, one complementary to a sequence on the (+) strand and the other to a downstream sequence on the (−) strand. Reiterative cycles of denaturation, annealing, and extension are used to generate multiple copies of the DNA that lies between the two primers. PCR has been used for a variety of purposes (7–13), including the detection of allelic polymorphisms and of DNA sequences unique to rare cells in a population, as well as the cross-species isolation of homologous genes. The feature common to all applications of PCR to date has been the use of primers designed to match two known or presumed genomic or cDNA sequences. In contrast, we have devised an application of this technique that achieves amplification and cloning of the region between a single short sequence in a cDNA molecule and its unknown 3′ or 5′ end. We demonstrate here the utility of this strategy, termed "rapid amplification of cDNA ends" (RACE), by using it to obtain clones representing transcripts of a gene, *int-2*, expressed at low abundance in the early mouse embryo.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

METHODS

3′-End Amplification of cDNAs. See Fig. 1 for the scheme. *Reverse transcription.* One microgram of poly(A)$^+$ RNA (ref. 14; ref. 15, pp. 91–98) in 16.5 μl of water was heated at 65°C for 3 min, quenched on ice, added to 2 μl of 10× RTC buffer (1× RTC buffer is 50 mM Tris·HCl, pH 8.15 at 41°C/6 mM $MgCl_2$/40 mM KCl/1 mM dithiothreitol/each dNTP at 1.5 mM), 0.25 μl (10 units) of RNasin (Promega Biotec, Madison, WI), 0.5 μl of $(dT)_{17}$-adaptor (1 μg/μl), and 10 units of avian myeloblastosis virus reverse transcriptase (Life Sciences, Saint Petersburg, FL), and incubated for 2 hr at 41°C. The reaction mixture was diluted to 1 ml with TE (10 mM Tris·HCl, pH 7.5/1 mM EDTA) and stored at 4°C.

Amplification. The cDNA pool (1 μl) and amplification (3′amp) and adaptor primers (25 pmol each) in 50 μl of PCR cocktail [10% (vol/vol) dimethyl sulfoxide/1× *Taq* polymerase buffer (New England Biolabs)/each dNTP at 1.5 mM] were denatured (5 min, 95°C) and cooled to 72°C. Then 2.5 units of *Thermus aquaticus* (*Taq*) DNA polymerase (Perkin–Elmer–Cetus) was added and the mixture was overlaid with 30 μl of mineral oil (Sigma) at 72°C and annealed at 50–58°C for 2 min. The cDNA was extended at 72°C for 40 min. Using a DNA Thermal Cycler (Perkin–Elmer–Cetus), we carried out 40 cycles of amplification by using a step program (94°C, 40 sec; 50–58°C, 2 min; 72°C, 3 min), followed by a 15-min final extension at 72°C.

5′-End Amplification of cDNAs. See Fig. 1. One microgram of poly(A)$^+$ RNA was reverse transcribed as described above except for the addition of 20 μCi (1 Ci = 37 GBq) of [^{32}P]dCTP and the substitution of 20 pmol of 5RT primer for $(dT)_{17}$-adaptor. Excess 5RT was removed as follows: the 20-μl cDNA pool was applied to a Bio-Gel A-5m (Bio-Rad) column (in a 2-ml serological pipette plugged with silane-treated glass wool) equilibrated with 0.05× TE. Void volume (0.8 ml) and 30 one-drop fractions were collected. Fractions −4 to +3 relative to the first peak of radioactivity were pooled, concentrated by centrifugation under reduced pressure (Speedvac), and adjusted to 23 μl. For tailing, 1 μl of 6 mM dATP, 6 μl of 5× tailing buffer (Bethesda Research Laboratories), and 15 units of terminal deoxynucleotidyl-transferase (Bethesda Research Laboratories) were added, and the mixture was incubated for 10 min at 37°C and heated for 15 min at 65°C. The reaction mixture was diluted to 500 μl in TE and 1- to 10-μl aliquots were used for amplification as described for the 3′-end procedure, except for the substitution of $(dT)_{17}$-adaptor (10 pmol), adaptor (25 pmol), and amplification (5′amp, 25 pmol) primers.

Southern and RNA Blot Analysis. Ten-microliter aliquots of RACE reaction products were separated by electrophoresis [1% agarose gel containing ethidium bromide (EtdBr) at 0.5

Abbreviations: RACE, rapid amplification of cDNA ends; PCR, (DNA) polymerase chain reaction; 3′amp and 5′amp, gene-specific primers for 3′- and 5′-end amplification; 5RT, gene-specific primer for reverse transcription; EtdBr, ethidium bromide.

Downloaded from https://www.pnas.org by 62.96.4.98 on September 28, 2023 from IP address 62.96.4.98.

Seikagaku Corporation Exhibit 2004 - Page 68 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

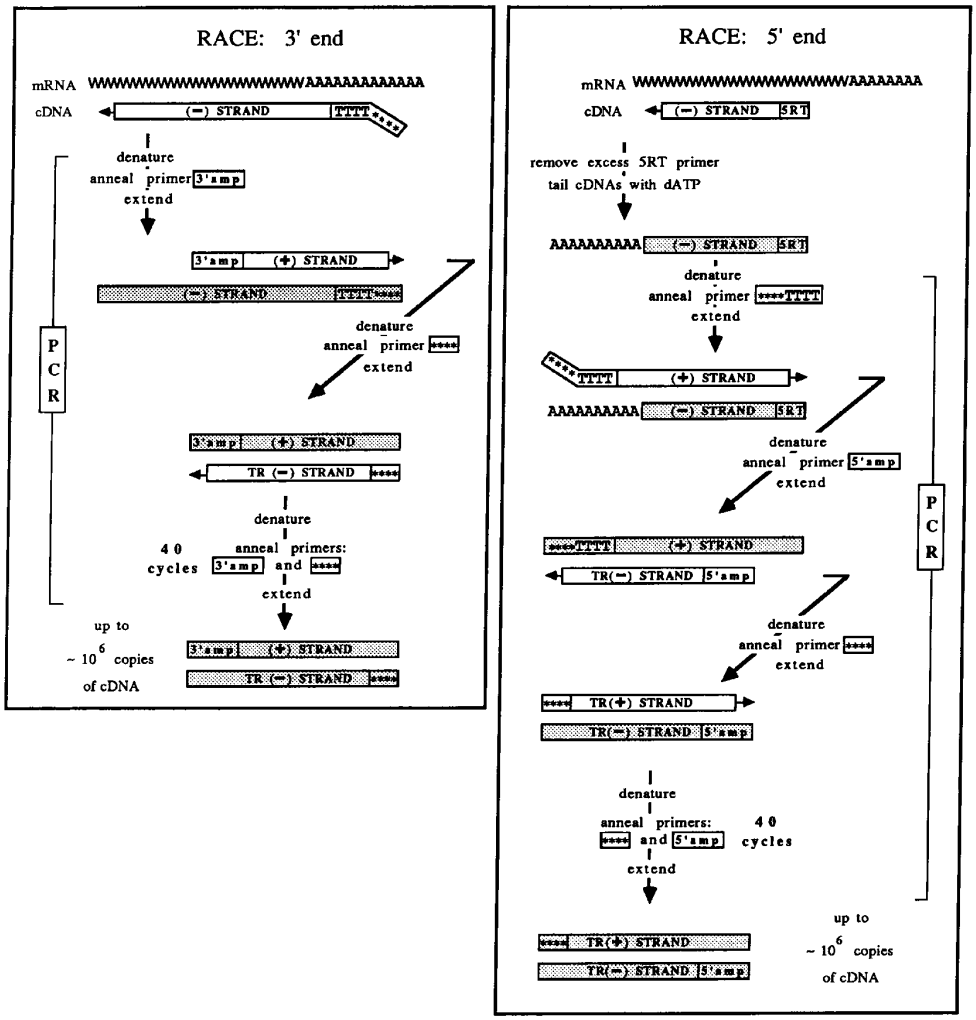


FIG. 1. Schematic representation of the RACE protocol. Primers: ****TTTT $(dT)_{17}$-adaptor, 5'-GACTCGAGTCGACATC-GATTTTTTTTTTTTTTTT-3'. This sequence contains the *Xho* I, *Sal* I, and *Cla* I recognition sites. ****, Adaptor, 5'-GACTCGAGTCGACATCG-3'. 3'amp (amp refers to amplification), specific to gene of interest, complementary to (−) strand. 5RT (RT refers to reverse transcription) and 5'amp, specific to gene of interest, complementary to (+) strand. Open rectangles represent DNA strands actively being synthesized; shaded rectangles represent DNA previously synthesized. At each step the diagram is simplified to illustrate only how the new product formed during the previous step is utilized. A (−) or (+) strand is designated as "truncated" (TR) when it is shorter than the original (−) or (+) strand, respectively.

μg/ml], transferred to GeneScreen (New England Nuclear; ref. 16), and hybridized at high stringency (17) with a ^{32}P-labeled probe (Bethesda Research Laboratories nick-translation kit). RNA blot analysis was carried out as previously described (17).

Cloning and Sequencing of cDNAs. RACE products were transferred into TE by using spun column chromatography (ref. 15, pp. 466–467), digested with restriction enzymes that recognize sites in the adaptor (*Cla* I or *Sal* I) or *int-2* sequences, and separated by electrophoresis. Regions of the gel containing specific products were isolated, and the DNA was extracted with Glassmilk (Bio 101, San Diego, CA) and cloned in a Bluescript vector (Stratagene, San Diego, CA). Plasmids with *int-2* cDNA inserts were identified by colony lift hybridization (ref. 15, p. 324). Restriction analyses (ref. 15, p. 104) were carried out on plasmid DNA prepared by the alkaline lysis method (18). Mini-prep plasmid DNA was sequenced with Sequenase (United States Biochemicals, Cleveland), using the supplier's recommendations.

It should be noted that whereas all 23 clones generated after *Sal* I cleavage of the adaptor contained an intact *Sal* I site, many of the clones generated after cleavage with *Cla* I were altered at the *Cla* I site and could no longer be cleaved by that enzyme (M.A.F., unpublished data). It is unknown whether such alteration was due to the positioning of the *Cla* I site immediately adjacent to the oligo(dT) sequence.

RESULTS

The RACE Protocol. To obtain cDNA 3' ends (Fig. 1), a RNA population is reverse transcribed to create a cDNA (−) strand. The primer used is a 35-base oligonucleotide with 17 dT residues and an adaptor sequence containing three endonuclease recognition sites that are found infrequently in genomic DNA [$(dT)_{17}$-adaptor; see legend to Fig. 1]. The

Downloaded from https://www.pnas.org by 62.96.4.98 on September 28, 2023 from IP address 62.96.4.98.

Seikagaku Corporation Exhibit 2004 - Page 69 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

presence of the adaptor places a unique sequence at the unknown end of the cDNA. Next, a gene-specific amplification primer (3'amp) is annealed with a small portion of the first (−) strand product (generally <1 ng) and extended to generate a complementary second (+) strand. Reiterative PCR cycles using 3'amp and adaptor primers then amplify the double-stranded cDNAs. Specificity of amplification is dependent on base-pairing of 3'amp only to molecules representing the mRNA of interest. The adaptor primer is used instead of the $(dT)_{17}$-adaptor primer because initial experiments suggested that long stretches of dT residues do not base pair well at the temperatures used to prevent mismatching of the specific primers.

A similar strategy is used to obtain cDNA 5' ends (Fig. 1). The initial specificity is achieved by using a gene-specific primer for reverse transcription (5RT). The primer-extended products are then separated from excess primer and a poly(A) tail is added by using dATP and terminal deoxynucleotidyltransferase. Second (+) strand synthesis is carried out with the $(dT)_{17}$-adaptor primer. Finally, the products are amplified by using the adaptor primer and a gene-specific amplification primer (5'amp) consisting of a sequence located upstream of the extension primer sequence. This "internal" primer increases the specificity and efficiency of the amplification reaction; it should bind only to the cDNAs of interest, whereas 5RT would bind to all (+) strand cDNAs present, including any generated through mismatched hybridization of 5RT during reverse transcription. Moreover, 5'amp would not bind to any residual 5RT that had been tailed. Such tailed 5RT, because of its small size and relatively large initial concentration, would serve as a potent substrate for a PCR reaction that included the 5RT and $(dT)_{17}$-adaptor primers, and this would greatly decrease the efficiency of the cDNA amplification.

After amplification, the products are analyzed by restriction and Southern blot analyses and cloned. The optimal cloning strategy utilizes one restriction endonuclease that cleaves within the adaptor sequence and a second that cleaves within the amplified region. This adds selectivity to the cloning process, since most of the nonspecific cDNA products will not be cleaved by the second enzyme and thus will not be cloned. Alternatively, the RACE products can be cloned intact by also incorporating a restriction site into the 5' end of the gene-specific primer. Finally, full-length cDNA clones can be reconstructed from separate but overlapping 3' and 5' RACE products, or they can be synthesized by using primers whose sequences were obtained by analysis of the extreme 5' and 3' ends of the RACE products for PCR amplification of reverse-transcribed mRNA.

Isolation of *int-2* cDNA Clones by Using the RACE Protocol. We tested this scheme by generating cDNA copies of mRNA from *int-2*, a gene characterized by multiple transcripts expressed at very low abundance (≈2 copies per cell, ref. 19), and whose complete sequence is known (20). Because we had recently isolated *int-2* cDNA clones by conventional methods (21), we had a standard against which we could judge the results of the present study. On the basis of our previous results, we expected to obtain *int-2* cDNA clones with alternative start and polyadenylylation sites (Fig. 2*A*).

The first 3'amp primer used, 3'amp-D, was specific to the long *int-2* mRNAs (Fig. 2*A*). The 3'-end cDNA product of the amplification reaction was expected to be ≈1050 bp in length. DNA of the expected size was visible after EtdBr staining (Fig. 2*B*, lane 3'D), and its identity was confirmed by Southern blot analysis using an *int-2* cDNA probe (Fig. 2*C*, lane 3'D). In addition, most samples contained *int-2* cDNAs of unexpected size, which were subsequently determined to be single-stranded molecules (see below). In some circumstances other reaction products were visible after EtdBr staining, but they did not hybridize to the *int-2* probe; we found that such nonspecific PCR products could be minimized by optimizing the annealing temperature (M.A.F., unpublished data). EtdBr-staining bands were also visible in the control sample, but they did not hybridize with the *int-2* probe (Fig. 2 *B* and *C*, lane 3'C).

The second 3'amp primer used, 3'amp-U, was common to both short and long classes (Fig. 2*A*) and thus was expected to generate cDNA products 480 and 1580 bp in length, respectively. Of these, only the 480-bp product was detectable by staining with EtdBr (Fig. 2*B*, lane 3'U), but the presence of both cDNAs was demonstrated by Southern blot analysis (Fig. 2*C*, lane 3'U). The observation that the 480-bp cDNA is ≈10-fold more abundant than the 1580-bp cDNA is inconsistent with the fact that the short classes of mRNA are only ≈2-fold more abundant than the long ones (Fig. 2*A*). This finding, that the relative abundance of different cDNA products generated by multiple cycles of PCR may not accurately reflect the composition of the mRNA population, is presumably due to a decrease in PCR efficiency as extension length increases.

When *int-2* cDNA 5' ends were produced by using the RACE strategy, we expected ≈690- and ≈930-bp products representing the 5' ends of the class B and the less abundant class A mRNAs, respectively (Fig. 2*A*). Although both fragments could be detected by Southern blot analysis, they were not always visible after EtdBr staining (Fig. 2 *B* and *C*, lane 5'P).

The specific products of the 3'- and 5'-end RACE reactions were isolated from agarose gels and cloned. The possibility afforded by the RACE protocol of separating different classes of cDNAs by gel electrophoresis prior to cloning was particularly useful in obtaining cDNAs representing class A mRNAs, which would otherwise have been difficult to detect in the presence of the much more numerous class B clones. However, we did not further size-select the RACE products in either class, and therefore we were able to examine the range of cDNAs produced by this protocol. Together the clones obtained represented all four transcripts expressed from the *int-2* gene, the longest of which is ≈2.9 kilobases (kb).

The cDNA clones representing the 3' ends of the long and short classes of *int-2* mRNA were expected to terminate at *int-2* genomic positions 7508 and 6404, respectively. Sequence analysis of four independent cDNA clones (three of the long and one of the short class) confirmed that the products obtained by the RACE method were identical to those obtained by conventional cDNA cloning methods (21). Previous studies suggested that the 5' ends of class A and B *int-2* mRNAs begin around positions 938 and 1677/78 of the *int-2* genomic sequence, respectively (21); however, single start sites are unlikely since the *int-2* promoters lack TATA boxes, and promoters of this type usually initiate transcription at multiple sites (23–25). Sequence analysis suggested that the majority of the clones studied were full length: 11/16 class B clones and 5/9 class A clones began at or within 21 bp of the presumed start sites (Fig. 2*D*). A higher frequency of full-length cDNA clones presumably could have been obtained by precise size selection of the RACE products prior to cloning.

In the course of developing and testing the RACE protocol, we discovered the following:

(*i*) For some mRNAs (not *int-2*) not all of the 3' end cDNAs were full length. We found this to be the case when we characterized cDNA clones of transcripts from the mouse *En-1* gene (17); sequence analysis of 3'-end *En-1* cDNA clones suggested that in some cases the $(dT)_{17}$-adaptor primer had bound to an A-rich region in the coding sequence (AAAGAAGAAAAAGAA) upstream of the poly(A) tail. This could be minimized by decreasing the concentration of the $(dT)_{17}$-adaptor primer to 15 ng/μg of mRNA (M.A.F., unpublished data).

(*ii*) Some of the specific reaction products appeared to be single-stranded cDNAs. These reaction products hybridized

Downloaded from https://www.pnas.org by 62.96.4.98 on September 28, 2023 from IP address 62.96.4.98.

Seikagaku Corporation Exhibit 2004 - Page 70 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

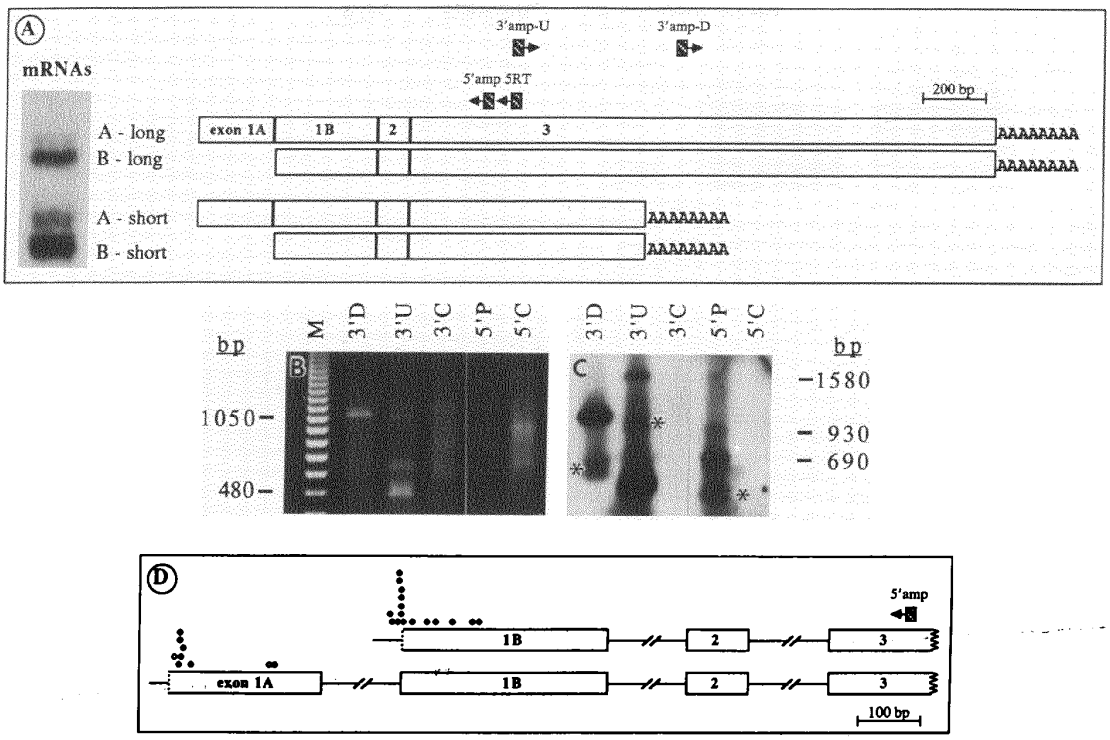


FIG. 2. Isolation of *int-2* cDNA clones by using the RACE protocol. (*A*) *int-2* mRNAs in teratocarcinoma-derived endodermal cells. (*Left*) RNA blot hybridized with an *int-2* exon 3 probe, showing four *int-2* transcripts (21). (*Right*) Depiction of structures of the four classes of *int-2* transcripts found in these cells as previously defined (21). Primers used are depicted as hatched boxes with arrows. Primer sequences: 3'amp primers 3'amp-U (5'-GGCAGAAGAAGCAGAGC-3') and 3'amp-D (5'AGAAGGCAGTGATCTGC-3'), complementary to (−) strand; 5RT (5'-GGCTCTGCTTCTTCTGC-3') and 5'amp primer (5'-CTCATGGTCCTTGTGGC-3'), complementary to (+) strand. bp, Base pairs. (*B* and *C*) Analysis of *int-2* RACE reaction products. Poly(A)$^+$ RNA isolated from endoderm-like cells obtained by treating F9 embryonal carcinoma cells with retinoic acid and dibutyryl-cAMP (22) was subjected to the RACE protocol. The reaction products were analyzed by electrophoresis and EtdBr staining (*B*) and Southern blot hybridization using an *int-2* cDNA probe (*C*). Marker lane (M) contains a 123-bp ladder (Bethesda Research Laboratories); sample lanes are labeled with the name of the gene-specific primer used in the amplification reaction (3'D, 3'amp-D, 3'U, 3'amp-U; 5'P, 5'amp) or as controls in which amplification was carried out in the absence of an *int-2* primer (3'C, 5'C). An * indicates single-stranded *int-2* cDNA. (*D*) Analysis of 5'-end *int-2* cDNA clones. The thin lines represent *int-2* genomic DNA and the open boxes the exons of the class A and B *int-2* mRNAs. The 5' ends of the mRNAs are symbolized by dotted vertical lines to indicate that multiple start sites are used. The jagged lines indicate that only the 5' portion of exon 3 is shown. The filled dots indicate the positions of the 5' ends of RACE clones as determined by sequence analysis, and the open dots the 5' ends as determined by restriction analysis.

to *int-2* probes, but they did not cleave with appropriate restriction endonucleases and were eliminated by digestion with mung bean nuclease (Fig. 3A). The quantity of these single-stranded products was also decreased by altering the buffer conditions (Fig. 3*A*, lane B), suggesting that complementary DNA strands are present and that appropriate binding can occur under suitable conditions. Single-stranded products are not usually observed with PCR, but use of a gene-specific primer in conjunction with a primer that binds to every cDNA present and that contains a long homopolymeric stretch may lead to greater production of one of the strands, or to mismatched pairing of some strands after the final round of PCR. Because these single-stranded products cannot easily be cloned, it is important to determine which of the specific products are double stranded.

(*iii*) The relative amount of longer cDNA products could be dramatically increased in several cases by substantially increasing the extension time of the initial (+) strand synthesis reaction (Fig. 3*B*). We also noted that small changes in the annealing temperatures (2°C) could cause significant changes in the ratio of specific to nonspecific cDNA amplification (M.A.F., unpublished data). In practice, we suggest using a conservative annealing temperature (50°C) for initial studies and raising it until the optimal temperature is reached for obtaining the cDNA product of interest, as judged by Southern blot analysis.

DISCUSSION

The RACE protocol described here provides an alternative to other cDNA cloning methods that is advantageous in many respects. A single gene-specific primer is used to generate 3'-end clones, whose sequences can be used, if necessary, to design the primers required for 5'-end cloning. The protocol involves few steps and the only reagents that must be specially prepared are the oligonucleotide primers. Moreover, steps common to other methods can be much simpler in the RACE protocol. For example, terminal deoxynucleotidyltransferase tailing conditions are generally optimized and monitored, since it is considered important to limit the length of the tail (4, 5). Using the RACE protocol, however, we found that the length of the poly(A) tail on the cDNA products did not exceed 34 nucleotides, even though no precautions were taken to limit tail length.

The RACE protocol is rapid and can be executed in 1 day. Moreover, it is unique in that substantial information about

Downloaded from https://www.pnas.org by 62.96.4.98 on September 28, 2023 from IP address 62.96.4.98.

Seikagaku Corporation Exhibit 2004 - Page 71 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

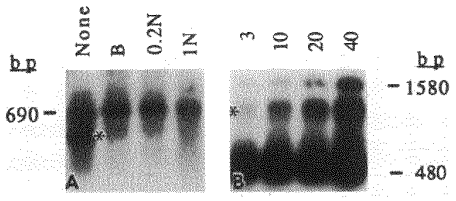


FIG. 3. (*A*) Sensitivity of specific RACE products to mung bean nuclease. *int-2* cDNAs were generated by using the RACE 5′-end protocol and transferred into TE by spun column chromatography. The samples were then incubated at 30°C for 30 min with no additions (None), with buffer only [B: 30 mM NaOAc, pH 5/50 mM NaCl/1 mM $ZnCl_2$/5% (vol/vol) glycerol], with buffer + 0.2 unit of mung bean nuclease (0.2N), or with buffer + 1 unit of mung bean nuclease (1N), and analyzed by Southern blot hybridization using an *int-2* probe. The 690-bp fragment is the amplified class B 5′-end cDNA; * indicates single-stranded *int-2* cDNA. (*B*) Effect of extension time on production of specific RACE products. *int-2* cDNA was prepared by using the RACE 3′ end protocol with 3′amp-U as described in *Materials and Methods*, except that the time allowed for the synthesis of the (+) strand during the first cycle of PCR was varied; in subsequent steps the extension time was 3 min. Samples were analyzed by Southern blotting. Lanes are designated according to the length of the extension time (3, 10, 20, and 40 min). The 480- and 1580-bp fragments are 3′-end cDNAs of the short and long *int-2* mRNAs, respectively; * indicates single-stranded *int-2* cDNA.

the cDNAs can be obtained prior to cloning and screening steps. Specifically, the RACE products can be examined by Southern blotting to determine whether the desired cDNAs are present and to evaluate the complexity of the cDNA population. This allows cDNAs representing unusual transcripts, such as those created by alternative splicing or by the use of alternative promoters, to be selected for cloning.

The RACE protocol is more efficient than other methods. This is evident from a comparison of the results of cloning *int-2* cDNAs by using conventional library construction vs. the RACE method. Starting with a total of 15 μg of poly(A)$^+$ RNA, we constructed two cDNA libraries, screened 2.4 × 10^6 recombinant phage, and isolated a total of 9 *int-2* cDNAs (21). In the present study we started with 0.5 μg of poly(A)$^+$ RNA, generated and screened several hundred recombinant plasmids with <0.05% of the RACE products, and isolated 29 independent *int-2* cDNA clones. Sequence analysis demonstrated that the 3′ and 5′ ends of all of four *int-2* mRNAs were accurately represented by the RACE products. Moreover, all of the 3′-end clones terminated at known polyadenylylation sites and the majority of the 5′-end RACE clones appeared to be full length. Thus, we have used this technique to readily obtain full-length cDNA clones representing transcripts 1.4 to 2.9 kb long. As our data suggest, however, PCR amplification becomes less efficient with increasing fragment length. Therefore, in cases in which the generation of full-length cDNAs would require amplification over a region more than several kilobases in length, production of the desired cDNAs may be difficult. Nevertheless, in most cases it should be possible to devise strategies to overcome this limitation. Another potential problem is that a limited number of errors are introduced by the use of multiple rounds of amplification (ref. 7; ≈0.1%, unpublished data of M.A.F.). Therefore, the sequence of a cDNA clone obtained by the RACE method should be verified by comparison with sequences of several independent cDNA clones or genomic DNA.

Other potential applications of this protocol are evident. For example, primers based on amino acid sequence might be adequate for RACE cDNA cloning, although our preliminary data suggest that the feasibility of this approach may be limited by nonspecific amplification due to primer mismatch. It should also be possible to use a modified RACE protocol to construct general cDNA libraries: one could reverse transcribe using $(dT)_{17}$-adaptor primer, tail the (−) strand products with G or C residues, generate a (+) strand with a different adaptor on its 5′ end, and amplify the pooled cDNAs. Since the RACE protocol as presently described requires less than 1 ng of mRNA (≈5000 copies of a low-abundance message) and this amount could be reduced considerably, it is conceivable that the construction of such general cDNA libraries ultimately could be carried out with the RNA from a single cell.

We are grateful to Monica Boyle and John Scott for technical assistance, and to our colleagues Frank Gertler, David J. Thornton, Carol Brenner, Barbara Levinson, Greg Barsh, Tom Cooper, and Tom Kornberg for helpful discussion and advice. We also thank Zena Werb for providing access to a DNA Thermal Cycler. M.A.F. and M.K.D. were supported by postdoctoral fellowships from the American Cancer Society and the Leukemia Society of America, respectively. The work was supported by Grant CD-319 from the American Cancer Society and Grant R01 HD20959 from the National Institutes of Health.

1. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
2. Heidecker, G. & Messing, J. (1987) *Methods Enzymol.* **159**, 28–40.
3. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. (1977) *Cell* **10**, 571–586.
4. Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokota, T. & Arai, K. (1987) *Methods Enzymol.* **159**, 3–27.
5. Coleclough, C. (1987) *Methods Enzymol.* **159**, 64–83.
6. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354.
7. Saiki, R. K., Gelfand, D. H., Stoffel, B., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
8. Kogan, S. C., Doherty, M. & Gitschier, J. (1987) *N. Engl. J. Med.* **317**, 985–990.
9. Embury, S. H., Scharf, S. J., Saiki, R. K., Gholson, M. A., Golbus, M., Arnheim, N. & Erlich, H. A. (1987) *N. Engl. J. Med.* **316**, 656–661.
10. Lee, M. S., Chang, K., Cabanillas, F., Freireich, E. J., Trujillo, J. M. & Stass, S. A. (1987) *Science* **237**, 175–178.
11. Lee, C. C., Wu, X., Gibbs, R. A., Cook, R. G., Muzny, D. M. & Caskey, C. T. (1988) *Science* **239**, 1288–1291.
12. Ou, C.-Y., Kwok, S., Mitchell, S. W., Mack, D. H., Sninsky, J. J., Krebs, J. W., Feorino, P., Warfield, D. & Schochetman, G. (1988) *Science* **239**, 295–297.
13. Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J. & Scott, J. (1987) *Cell* **50**, 831–840.
14. Cathala, G., Savouret, J., Mendez, B., West, B. L., Karin, M., Martial, J. A. & Baxter, J. D. (1983) *DNA* **2**, 329–335.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
17. Joyner, A. L., Kornberg, T., Coleman, K., Cox, D. & Martin, G. R. (1985) *Cell* **43**, 29–37.
18. Birnboim, C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
19. Jakobovits, A., Shackleford, G. M., Varmus, H. E. & Martin, G. R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7806–7810.
20. Moore, R., Casey, G., Brookes, S., Dixon, M., Peters, G. & Dickson, C. (1986) *EMBO J.* **5**, 919–924.
21. Mansour, S. B. & Martin, G. R. (1988) *EMBO J.* **7**, 2035–2041.
22. Strickland, S., Smith, K. K. & Marotti, K. R. (1980) *Cell* **21**, 347–355.
23. Melton, D. A., Konecki, D. S., Brennand, J. & Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2147–2151.
24. Reynolds, G. A., Baser, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L. & Luskey, K. L. (1984) *Cell* **38**, 275–285.
25. Ishii, S., Merlino, G. T. & Pastan, I. (1985) *Science* **230**, 1378–1381.

Downloaded from https://www.pnas.org by 62.96.4.98 on September 28, 2023 from IP address 62.96.4.98.

Seikagaku Corporation Exhibit 2004 - Page 72 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

cDNA Amplification Using One-Sided (Anchored) PCR

Conventional approaches to the characterization and sequencing of mRNA generally rely on the preparation and screening of a cDNA library, and on the subsequent isolation of single cDNA clones from within the library (*UNITS 5.5, 5.6 & 5.8A*). While this methodology has proven useful overall, success depends critically on the quality of the cDNA library. Frequently, cDNA libraries are vitiated by prematurely terminated cDNA molecules that do not reflect full-length message. In addition, rare messages may frequently be lost altogether in the preparation of cDNA libraries.

The polymerase chain reaction (PCR) offers an opportunity to directly select, amplify, and isolate a message of interest. This protocol presents a modification of PCR, called anchored PCR, that allows amplification of full-length mRNA when only a small amount of sequence information is available. Unlike the amplification of RNA or cDNA by conventional PCR (*UNIT 15.4*)—which requires prior knowledge of the sequences flanking the region of interest to design the PCR primers—anchored PCR can be employed when only a small region of sequence lying within the mRNA is known in advance.

In theory, only a single specific primer is required to perform anchored PCR. In practice, amplification of a single product requires a second round of PCR amplification, using a second sequence-specific primer. This scheme is depicted in Figures 15.6.1 and 15.6.2. Both the original and reamplifications use an oligo(dT) primer complementary either to the poly(A) tail of the mature mRNA [when amplifying downstream (3′) to the known sequence] or to an enzymatically synthesized homopolymer tail added to the cDNA following first strand synthesis [when amplifying upstream (5′) to the known sequence]. The two rounds of PCR amplification result in a single product that can be sequenced directly or cloned into an appropriate vector for further analysis.

This unit presents two basic protocols for the amplification of cDNA targets using anchored PCR. The first basic protocol can be used to amplify targets where the downstream (3′) sequences are unknown. The second basic protocol permits amplification of target molecules whose upstream (5′) sequences are unknown.

AMPLIFICATION OF REGIONS DOWNSTREAM (3′) OF KNOWN SEQUENCE

BASIC PROTOCOL

This method can be used to characterize messages of interest from any source of poly(A)$^+$ RNA prepared by conventional protocols (*UNITS 4.1-4.5*). Although the amount of poly(A)$^+$ RNA required will depend on the genome complexity of the organism being studied and on the relative abundance of the targeted mRNA, in general 100 to 300 ng of total poly(A)$^+$ RNA prepared from vertebrate tissue is sufficient for this procedure. Three PCR primers are required: a 20-mer oligo(dT) primer and two sequence-specific primers, one for the original amplification and the second for the reamplification (see Fig. 15.6.1). The internal sequence-specific primer can be immediately adjacent (3′) to or can partially overlap primer 1 at the 3′ end.

Materials

Source of RNA (*UNITS 4.1-4.5*)
5× Moloney murine leukemia virus (MoMLV) reverse transcriptase buffer
5 µg/µl bovine serum albumin (BSA)
10 mM 4dNTP mix (10 mM each dNTP in TE buffer, pH 7.5, stored at –20°C)
500 ng/µl actinomycin D
200 U/µl MoMLV reverse transcriptase (Table 3.4.1 & *UNIT 3.7*)

Contributed by Robert L. Dorit and Osamu Ohara
Current Protocols in Molecular Biology (1992) 15.6.1-15.6.10
Copyright © 2000 by John Wiley & Sons, Inc.

Seikagaku Corporation Exhibit 2004 - Page 73 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

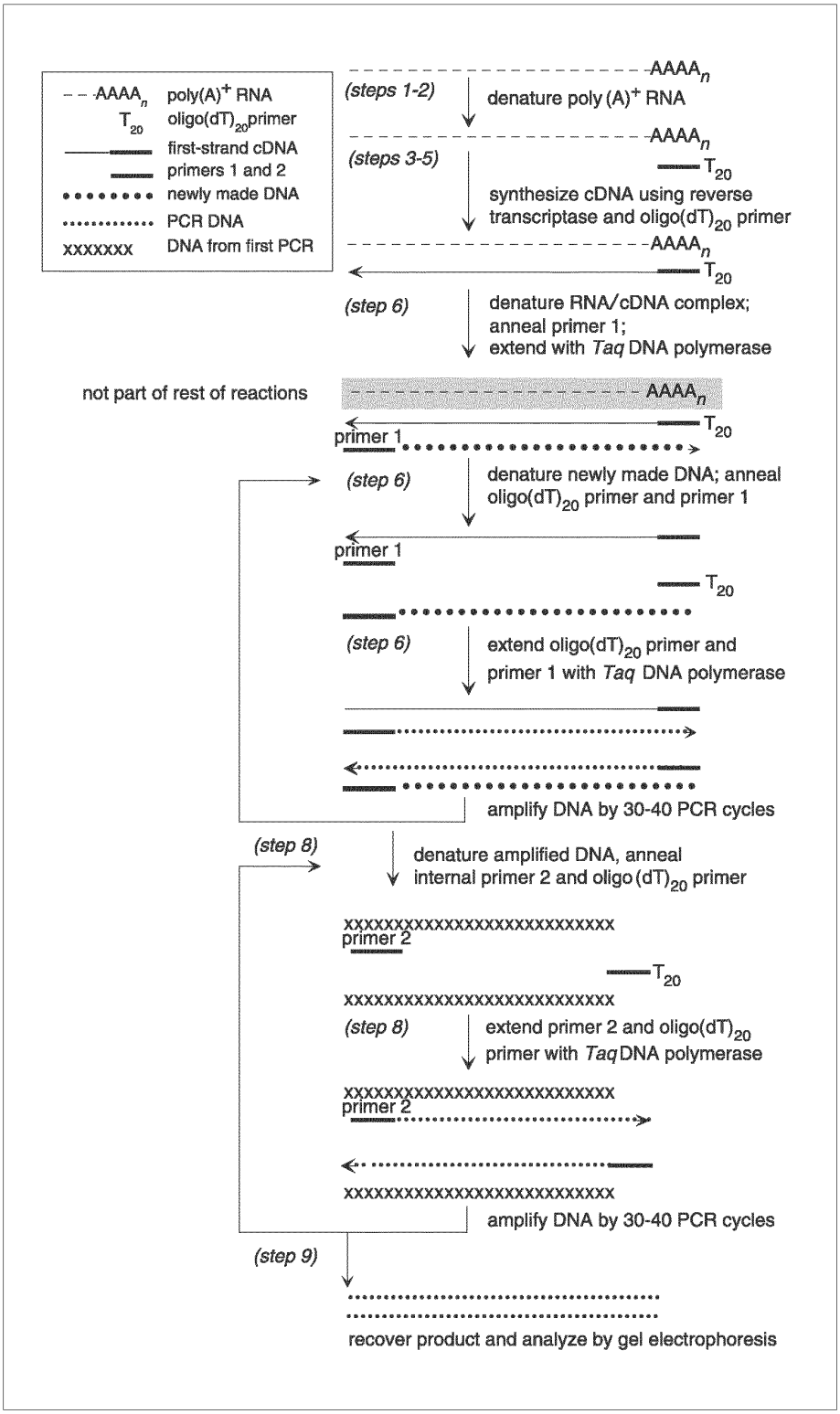


Figure 15.6.1 Flow chart of downstream (3′) anchored PCR protocol (see text for details). The steps correspond to those listed in the basic protocol. Note that primer 2 can be immediately adjacent (3′) to, or can partially overlap with, primer 1.

Seikagaku Corporation Exhibit 2004 - Page 74 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

15 pmol/µl (100 ng/µl) oligo(dT)$_{20}$ primer
100 pmol/µl each of sequence-specific primers 1 and 2 (see Fig. 15.6.1)
TE buffer (*APPENDIX 2*)
2.5 mM 4dNTP mix (2.5 mM each dNTP in TE buffer, pH 7.5, stored at –20°C)
10× amplification buffer
2.5 U/µl *Taq* DNA polymerase (*UNITS 15.1 & 3.5*)
Mineral oil

Additional reagents and equipment for preparation of poly(A)$^+$ RNA (*UNIT 4.1*), PCR amplification (*UNIT 15.1*), agarose gel electrophoresis (*UNIT 2.5*), and Southern blotting and hybridization (*UNITS 2.9 & 6.4*)

1. Prepare 100 ng poly(A)$^+$ RNA at a concentration of ≥100 ng/µl and place in a 1.5-ml microcentrifuge tube.

   This high concentration is necessary because the final volume of the reverse transcriptase reaction should be kept as low as possible, preferably ≤10 µl. Larger amounts of poly(A)$^+$ RNA (up to 1 µg) can be used if readily available. Amounts below 100 ng may not include the full complement of rare mRNAs.

2. Incubate poly(A)$^+$ RNA 2 min at 65°C. Microcentrifuge briefly at room temperature and place immediately on ice.

   This incubation melts the secondary structure of the mRNA, removing hairpins and loops that interfere with the synthesis of cDNA.

3. In a clean 1.5-ml microcentrifuge tube, prepare the following reverse transcriptase reaction mix on ice (10 µl total):

   2 µl 5× MoMLV reverse transcriptase buffer
   1 µl (5 µg) BSA
   1 µl poly(A)$^+$ RNA (from step 2)
   1 µl 10 mM 4dNTP mix
   1 µl (500 ng) actinomycin D
   1 µl (15 pmol) oligo(dT)$_{20}$ primer
   2 µl sterile H_2O
   1 µl MoMLV reverse transcriptase.

4. Mix gently, microcentrifuge briefly at room temperature and incubate 1 hr at 37°C.

   This results in synthesis of cDNA.

5. Add 40 µl TE buffer to bring the total reaction volume up to 50 µl.

6. Prepare the following PCR reaction mix on ice (100 µl final):

   1 µl cDNA template (from step 5)
   10 µl 10× amplification buffer
   1 µl (100 pmol) oligo(dT)$_{20}$ primer
   1 µl (100 pmol) sequence-specific primer 1
   6 µl 2.5 mM 4dNTP mix
   80 µl sterile H_2O
   1 µl (2.5 U) *Taq* DNA polymerase.

   Overlay reaction mixture with mineral oil. Carry out 30 to 40 cycles of amplification (*UNIT 15.1*).

   Initially, amplifications should be carried out in the presence of 1.5 mM $MgCl_2$ (final concentration), with annealing temperatures not exceeding 42°C. These conditions can be subsequently modified and optimized for the specific template used (UNIT 15.1). Primer

Seikagaku Corporation Exhibit 2004 - Page 75 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

concentrations can be reduced to 30 pmol of each primer without reducing the efficiency of amplification.

7. Analyze an aliquot by agarose gel electrophoresis.

   The result of this first PCR is usually a smear around the expected size range, because specificity is seldom conferred by a single sequence-specific primer. The presence of the desired product can confirmed by hybridizing a Southern blot (UNITS 2.9 & 6.4) of the gel containing the PCR products to the sequence-specific primer 2 internal to the amplified sequence.

8. Remove a 1-µl aliquot of the amplification product from step 6 to serve as the template for reamplification. Carry out a second round of PCR as described in step 6, using 40 to 100 pmol each of the oligo(dT)$_{20}$ primer and the internal sequence-specific primer 2.

   Although 100 pmol of each primer are usually added in conventional PCR amplifications, lower concentrations of primer do not appear to reduce amplification efficiency or yield, and may reduce the cost of this procedure.

9. Analyze an aliquot of the second amplification by agarose gel electrophoresis. The amplified product should now appear as a single band on an agarose gel.

10. If desired, characterize the amplification product by cloning into an appropriate vector (*UNIT 3.16*) or by direct sequencing (*UNIT 15.2*).

BASIC PROTOCOL

AMPLIFICATION OF REGIONS UPSTREAM (5′) OF KNOWN SEQUENCE

In contrast to the first protocol, this procedure uses one of the sequence-specific primers to initiate the synthesis of the cDNA strand. This cDNA is modified by the addition of a poly(A) tail. PCR amplifications—mediated by two sequence-specific primers and a 20-mer oligo(dT) primer complementary to the newly synthesized tail—yield the desired unique product (Fig. 15.6.2).

Materials

Source of RNA (*UNITS 4.1-4.5*)
100 pmol/µl sequence-specific primers 3 and 4 (Fig. 15.6.2)
15 pmol/µl (~100 ng/µl) oligo(dT)$_{20}$ primer
1 M NaCl
200 mM Tris·Cl, pH 7.5 (*APPENDIX 2*)
25 mM EDTA
100% and 70% ethanol, ice-cold
5× Moloney murine leukemia virus (MoMLV) reverse transcriptase buffer
5 µg/µl bovine serum albumin (BSA)
10 mM 4dNTP mix (10 mM each dNTP in TE buffer, pH 7.5, stored at –20°C)
500 ng/µl actinomycin D
200 U/µl MoMLV reverse transcriptase (Table 3.4.1 & *UNIT 3.7*)
3 M sodium acetate (*APPENDIX 2*)
TE buffer, pH 7.5 (*APPENDIX 2*)
5× terminal deoxynucleotidyltransferase (TdT) buffer
15 mM $CoCl_2$
1 mM dATP (*UNIT 3.4*)
Terminal transferase (Table 3.4.1 & *UNIT 3.6*)
40°C water bath

Additional reagents and equipment for preparation of poly(A)$^+$ RNA (*UNIT 4.5*), phenol extraction (*UNIT 2.1*), and PCR amplification (*UNIT 15.1*)

Seikagaku Corporation Exhibit 2004 - Page 76 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

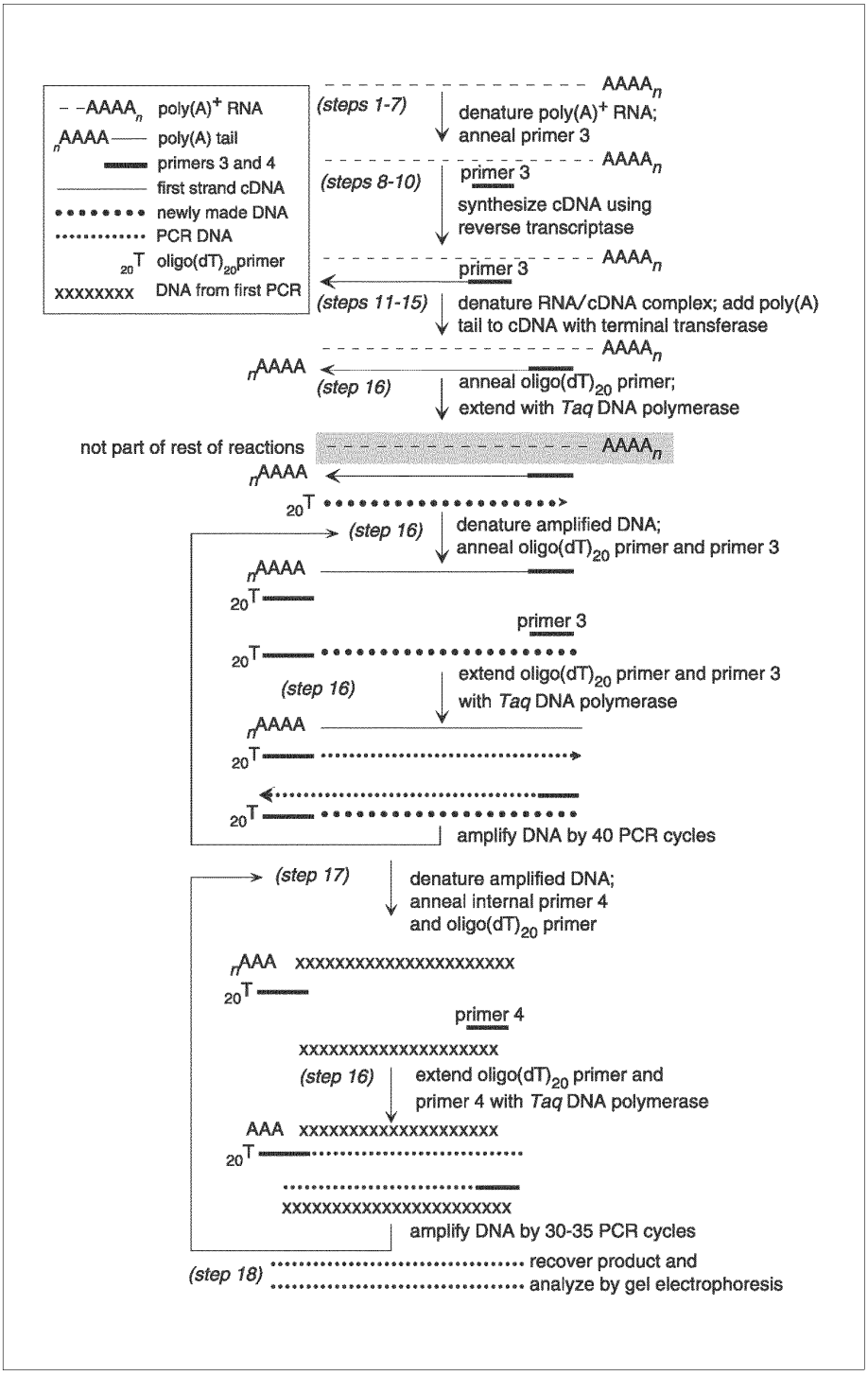


Figure 15.6.2 Flow chart of upstream (5′) anchored PCR protocol (see text for details). The steps correspond to those listed in the protocol. Note that primer 4 can be immediately adjacent to, or can partially overlap with, primer 3.

Seikagaku Corporation Exhibit 2004 - Page 77 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Prepare RNA and anneal sequence-specific primer

1. Prepare 100 ng of poly(A)$^+$ RNA at a concentration of 100 ng/µl.
2. Prepare a 5-µl annealing mix containing:

   1 µl (1 *fmol*) sequence-specific primer 3
   1 µl 1 M NaCl
   1 µl 200 mM Tris·Cl, pH 7.5
   1 µl 25 mM EDTA
   1 µl (100 ng) poly(A)$^+$ RNA (from step 1).

   Amount of primer to use for sequence-specific cDNA synthesis depends in part on the abundance of target mRNA. As a rule, a 5- to 10-fold molar excess of oligomer primer relative to target template yields the best results. Note that small amount of primer boosts specificity of targeted cDNA synthesis. Primer 3 should be diluted with water or 0.1× TE buffer, pH 7.5, to reach the desired concentration.
3. Incubate annealing mix 3 min at 65°C, microcentrifuge briefly at room temperature, and place immediately on ice for 2 min.

   This incubation melts the secondary structure of the mRNA, removing hairpins and loops that interfere with the synthesis of cDNA. "Snap-cooling" on ice prevents the secondary structure from re-forming.
4. Incubate 3 to 4 hr at 40°C.

   The long incubation time and low primer concentration increase the specificity of the primer-template annealing.
5. Add 15 µl ice-cold 100% ethanol, place in dry ice/ethanol bath for 10 min, and microcentrifuge 10 min at high speed, 4°C.
6. Decant supernatant. Add 50 µl ice-cold 70% ethanol and gently invert tube several times to rinse and desalt the pellet (do not vortex). Microcentrifuge 2 min at room temperature, remove supernatant, and dry briefly under vacuum.
7. Resuspend pellet in 10 µl water.

Synthesize first-strand cDNA

8. On ice, prepare reverse transcriptase mix containing (25 µl total):

   5 µl 5× MoMLV reverse transcriptase buffer
   2.5 µl (12.5 µg) BSA
   2.5 µl 10 mM 4dNTP mix
   2.5 µl (1.25 µg) actinomycin D
   10 µl annealed primer/template (from step 7)
   1.5 µl sterile H_2O
   1 µl (200 U) MoMLV reverse transcriptase.

   Incubate 1 hr at 37°C.
9. Phenol extract the sample. Carefully transfer supernatant to new microcentrifuge tube. Add 2.5 µl of 3 M sodium acetate (0.3 M final) and 75 µl ice-cold 100% ethanol. Place 5 min in dry ice/ethanol bath, then microcentrifuge 20 min at high speed, 4°C. Discard supernatant.

   In cases where small amounts of poly(A)$^+$ RNA are being used, 1 to 5 µg carrier tRNA (UNIT 4.6) may be added at this step to facilitate the precipitation.
10. Resuspend pellet in 25 µl of TE buffer, pH 7.5. Add 2.5 µl of 3 M sodium acetate (0.3 M final) and repeat the ethanol precipitation as described in step 9, discarding supernatant after spin.

Seikagaku Corporation Exhibit 2004 - Page 78 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

11. Add 100 µl ice-cold 70% ethanol to pellet and rinse by gently inverting tube several times (do not vortex). Microcentrifuge 5 min at high speed, room temperature; discard supernatant, and dry pellet briefly under vacuum.

12. Resuspend pellet in 5 µl water and boil 2 min. Microcentrifuge briefly at room temperature and place immediately on ice.

    This step denatures the cDNA/RNA hybrid molecules prior to tailing the cDNA molecule with terminal transferase.

Add poly(A) tail to cDNA

13. On ice prepare terminal transferase mix in order listed below (10 µl final):

    2 µl 5× TdT buffer
    1 µl 15 mM $CoCl_2$ (1.5 mM final)
    1 µl 1 mM dATP (100 µmol final)
    5 µl cDNA mix (from step 12)
    1 µl (25 U) terminal transferase.

    Incubate 30 min at 37°C.

14. Inactivate enzyme by heating 2 min at 65°C.

15. Add 1 µl of 3 M sodium acetate. Ethanol precipitate, wash and dry pellet as described in step 10.

Amplify by PCR and analyze results

16. Resuspend pellet in 10 µl water. Carry out 40 amplification cycles as in step 6 of the first basic protocol, using 40 to 100 pmol each of sequence-specific primer 3 and oligo$(dT)_{20}$.

    Although 100 pmol of each primer are usually added in conventional PCR amplifications, lower concentrations of primer do not appear to reduce amplification efficiency or yield, and may reduce the cost of this procedure.

    Initially, amplifications should be carried out in the presence of 1.5 mM $MgCl_2$ (final), with annealing temperatures not exceeding 42°C. These conditions can be subsequently modified and optimized for the specific template used (UNIT 15.1).

    The products of this round of PCR will not be visualized by ethidium bromide staining of an agarose gel, but can be seen when a Southern blot of the gel is hybridized to sequence-specific primer 4 internal to the amplified region.

17. Remove a 1-µl aliquot of the amplification product of step 16 to serve as template for a new round of PCR amplification. Carry out 35 cycles of PCR as in step 6 of the first basic protocol, using 40 to 100 pmol each of primer 4 and the oligo(dT) primer.

    Primer 4 can be immediately adjacent to, or partially overlapping with, primer 3. Because the anchored PCR procedures are carried out independently in the 3′ and 5′ directions, sequence-specific primers 3 and 4 can be the complements of primers 1 and 2 used in the previous protocol (see Fig. 15.6.2).

18. Analyze a 10-µl aliquot by agarose gel electrophoresis.

    The amplified product, stretching from the target site of the sequence-specific primer 4 to the 5′ end of the cDNA, should now appear as a single band on an agarose gel.

19. If desired, characterize the PCR product by cloning into an appropriate vector (*UNIT 3.16*) and/or by direct sequencing (*UNIT 15.2*).

Seikagaku Corporation Exhibit 2004 - Page 79 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

REAGENTS AND SOLUTIONS

5× MoMLV reverse transcriptase buffer

250 mM Tris·Cl, pH 8.3
375 mM KCl
50 mM DTT
15 mM $MgCl_2$
Store at –20°C for ≤3 months

10× amplification buffer

500 mM KCl
100 mM Tris·Cl, pH 8.8
15 mM $MgCl_2$
30 mM DTT
1 mg/ml BSA
Store at –20°C for ≤3 months

5× terminal transferase (TdT) buffer

1 M potassium cacodylate
125 mM Tris·Cl, pH 7.4
1.25 μg/μl BSA
Store at –20°C for ≤6 weeks

COMMENTARY

Background Information

The advent of PCR has spawned a number of procedures that do not require full knowledge of the sequence being amplified (Frohman et al., 1988; Ochman et al., 1988; Loh et al., 1989; Mueller and Wold, 1989; Ohara et al., 1989; Tam et al., 1989), and that assume only a small region of sequence is known. In many of these methods, the PCR amplifications are anchored by a primer derived from the region of known sequence; the second primer is nonspecific and is targeted to a general feature of the sequence, such as the poly(A) tail present on most mature mRNAs. Alternatively, the target sequence can be enzymatically modified to provide an annealing site for the second primer. Such modifications include the addition (via terminal transferase) of a homopolymeric tail (*UNIT 3.6*) or the ligation of a sequence "splint" at the end of the target molecule (*UNIT 15.5*).

The protocols presented here are specifically designed for the amplification of full-length cDNA when only a fragment of the cDNA sequence is known. The amplification into the unknown downstream (3′) region capitalizes on the poly(A) tail present on most mRNAs. It therefore will not succeed with poly(A)$^-$ mRNAs, such as those encoding histones. The region of interest is amplified by annealing the oligo(dT) primer to this poly(A) tail and the sequence-specific primer to the appropriate known region of the cDNA.

Amplification into the unknown upstream (5′) region makes use of an artificial poly(A) tail that is enzymatically added to the 3′ end of an initial, specific cDNA product. In theory, then, upstream and downstream amplifications can succeed with any mRNA for which sufficient sequence to design a single sequence-specific oligonucleotide is known.

In practice, however, a single sequence-specific primer does not always confer sufficient selectivity in the amplification, and the primer may bind to inappropriate or partially complementary sequences, resulting in the amplification of more than one sequence. This problem is further compounded in situations where the only sequence available at the outset comes from the partial sequencing of the protein produced by the mRNA of interest, or from the identification of conserved regions in homologous mRNAs extracted from other organisms. Such cases require the use of degenerate sequence-specific primers, further reducing the specificity of the anchored PCR amplifications. The consequences of using degenerate primers will depend on the template being amplified, the degree of degeneracy, and the amplification conditions. The use of nondegenerate sequence-specific primers is strongly recommended for the protocols described here (Ohara et al., 1989).

Seikagaku Corporation Exhibit 2004 - Page 80 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Critical Parameters and Troubleshooting

As with any method for the analysis of mRNA, the success of these protocols depends on the quality and integrity of the mRNA being used. Appropriate precautions to ensure high yield of intact mRNA are detailed in Chapter 4 (see introduction to that chapter). Although these protocols can be performed using untreated microcentrifuge tubes and pipet tips, several steps involve single-stranded nucleic acids (mRNA, first-strand cDNA) and the use of silanized microcentrifuge tubes and pipet tips (*APPENDIX 3*) may be desirable.

It is important to keep in mind that the initial stages of anchored PCR will seldom result in a fully specific amplification, and the amplification products will most often appear as a smear when visualized on an agarose gel. The PCR products produced by this protocol should be monitored by Southern blot hybridization to a target-specific internal oligomer. Such monitoring confirms that the desired product is in fact being amplified in the initial phases of the protocols.

A number of steps may be taken to increase the target specificity of anchored PCR. As described in these protocols, a reamplification carried out using a second internal (nested) sequence-specific primer greatly reduces spurious amplification products. This second primer can lie immediately adjacent to the first specific primer or can partially overlap at the 3′ end. Thus, 40 to 50 nucleotides of known sequence are sufficient to fully anchor this two-step procedure.

A second problem that may interfere with the production of a single, distinct PCR product stems from the length of either the naturally occurring or the enzymatically synthesized poly(A) tail. A long poly(A) tail provides a large number of potential pairing sites for the nonspecific oligo(dT) primer. The resultant PCR products may consequently be of variable length.

The difficulties that stem from the variable length of the homopolymeric tails can be partially resolved by synthesizing an oligo(dT) primer with an additional specific sequence at the 5′ end, as is done in the race amplification of cDNA ends (RACE) protocol (Frohman et al., 1988). Subsequent reamplifications will then rely on a primer directed to this synthetic sequence. Alternatively, the oligo(dT) primer can be synthesized with a degenerate (A,G,C) 3′-most position, resulting in more precise annealing at the poly(A) tail boundary of the mRNA.

Finally, specificity increases if the first-round amplification products are size-selected by extracting an aliquot of the main PCR band within the smear (or of the "correct" target band, previously determined by Southern hybridization with an internal specific oligomer). The selected products can serve as templates for a second round of PCR amplification. Thus, the first-round products are run out on a low gelling/melting agarose gel (*UNIT 2.6*), and the desired target is size selected by piercing the appropriate band on the gel with a Pasteur pipet. This small gel core sample can then be placed directly into a new PCR mix and reamplified without further purification.

The products of this second amplification may still appear as a smear when visualized by ethidium bromide staining of an agarose gel (see commentary, *UNIT 15.1*). This smear results from the multiple positions that the oligo$(dT)_{20}$ primer can occupy along the length of the poly(A) tail. In such cases, select a small aliquot of product at the upper range of the smear and reamplify with 20 to 25 PCR cycles using the oligo$(dT)_{20}$ and internal primers (primers 2 or 4).

We have found that certain DNA targets cannot be efficiently amplified using standard PCR conditions. In such cases, low-yield PCR products can be readily cloned into an appropriate vector (Holton and Graham, 1991; Marchuck et al., 1991) and subsequently screened by colony hybridization (*UNIT 6.2*) to isolate the appropriate clones. Given the relatively high error rate of *Taq* polymerase (*UNITS 3.5 & 15.1*), sequences obtained from cloned PCR fragments should be confirmed by isolating and sequencing multiple clones.

Time Considerations

With the mRNA already in hand, each of the basic protocols will require 1.5 to 2 days to complete. Upstream and downstream anchored PCR can be carried out synchronously.

Literature Cited

Frohman, M.A., Dush, M.K., Martin, G.R. 1988. Rapid production of full length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Nat. Acad. Sci U.S.A.* 85:8998-9002.

Holton, T.A. and Graham, M.W. 1991. A simple and direct method for direct cloning of PCR products using ddT-tailed vectors. *Nucl. Acids Res.* 19:1156.

Seikagaku Corporation Exhibit 2004 - Page 81 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Loh, E.Y., Elliott, J.F., Cwirla, S., Lanier, L.L., and Davis, M.M. 1989. Polymerase chain reaction with single-sided specificity: Analysis of a T-cell receptor delta chain. *Science* 243:217-220.

Marchuck, D., Drumm, M., Saulino, A., and Collins, F.S. 1991. Construction of T-vectors: A rapid and general system for direct cloning of unmodified PCR products. *Nucl. Acids Res.* 19:1154.

Mueller, P.R. and Wold, B. 1989. In vivo footprinting of a muscle-specific enhancer by ligation-mediated PCR. *Science* 246: 780-786.

Ochman, H., Gerber, A.S., and Hartl, D.L. 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics* 120:621-623.

Ohara, O., Dorit, R.L. and Gilbert, W. 1989. One-sided polymerase chain reaction: The amplification of cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 86:5673-5677.

Tam, A.W., Smith, M.M., Fry, K.E., and Larrick, J.W. 1989. Construction of cDNA libraries from small numbers of cells using sequence independent primers. *Nucl. Acids Res.* 17: 1269.

Key Reference

Ohara et al., 1989. See above.

The protocols described here are used to isolate and characterize α-tropomyosin transcripts in common frog and zebra fish.

Contributed by Robert L. Dorit
Yale University
New Haven, Connecticut

Osamu Ohara
Shionogi Research Laboratories
Osaka, Japan

Seikagaku Corporation Exhibit 2004 - Page 82 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

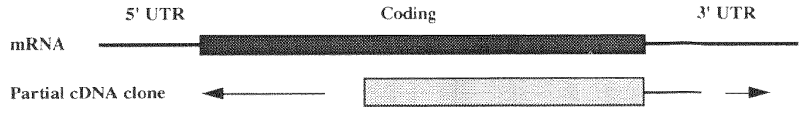
Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

On Beyond Classic RACE (Rapid Amplification of cDNA Ends)

Michael A. Frohman

Department of Pharmacology, University Medical Center at Stony Brook, Stony Brook, New York 11794-8651

Most attempts to identify and isolate a novel cDNA result in the acquisition of clones that represent only a part of the mRNA's complete sequence. The missing sequence (cDNA ends) can be cloned by PCR, using a technique variously called rapid amplification of cDNA ends (RACE),[1] anchored PCR,[2] or one-sided PCR.[3] Since the initial reports of this technique, many laboratories have developed significant improvements on the basic approach.[4–16] I will describe here the most recent hybrid version of the relatively simple classic RACE, and a more powerful but technically more challenging new RACE protocol that is adapted from the work of a number of laboratories.[17–24] Commercial RACE kits that are convenient but not as powerful as the most recent versions of classic and new RACE are available from Bethesda Research Laboratories[11] and Clontech.



Shown is a schematic representation of the setting in which RACE is useful in cDNA cloning strategies. Depicted is an mRNA for which a cDNA representing only an internal portion of the transcript has been obtained. Such circumstances often arise; one such example is when closely related genes are cloned using PCR amplification with degenerate primers encoding sequences homologous to amino acids found in all known members of the gene family.

OVERVIEW

Why use PCR (RACE) at all instead of screening (additional) cDNA libraries? RACE cloning is advantageous for several reasons: First, it takes weeks to screen cDNA libraries, obtain individual cDNA clones, and analyze the clones to determine whether the missing sequence is present. Using PCR, such information can be generated within a few days. As a result, it becomes practical to modify RNA preparation and/or reverse transcription conditions until full-length cDNAs are generated and observed. In addition, essentially unlimited numbers of independent clones can be generated using RACE, unlike library screens in which generally a single to a few cDNA clones are recovered. The availability of large numbers of clones provides confirmation of nucleotide sequence and allows the isolation of unusual transcripts that are alternately spliced or that begin at infrequently used promoters.

PRINCIPLES

Classic RACE

PCR is used to amplify partial cDNAs representing the region between a single point in a mRNA transcript and its 3′ or 5′ end (Fig. 1). A short internal stretch of sequence must already be known from the mRNA of interest. From this sequence, gene-specific primers are chosen that are oriented in the direction of the missing sequence. Extension of the partial cDNAs from the unknown end of the message back to the known region is achieved using primers that anneal to the preexisting poly(A) tail (3′ end) or an appended homopolymer tail (5′ end). Using RACE, enrichments on the order of 10^6 to 10^7-fold can be obtained. As a result, relatively pure cDNA "ends" are generated that can be easily cloned or rapidly characterized using conventional techniques.[1]

4:S40–S58©1994 by Cold Spring Harbor Laboratory ISSN 1054-9805/94 $1.00

Seikagaku Corporation Exhibit 2004 - Page 83 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

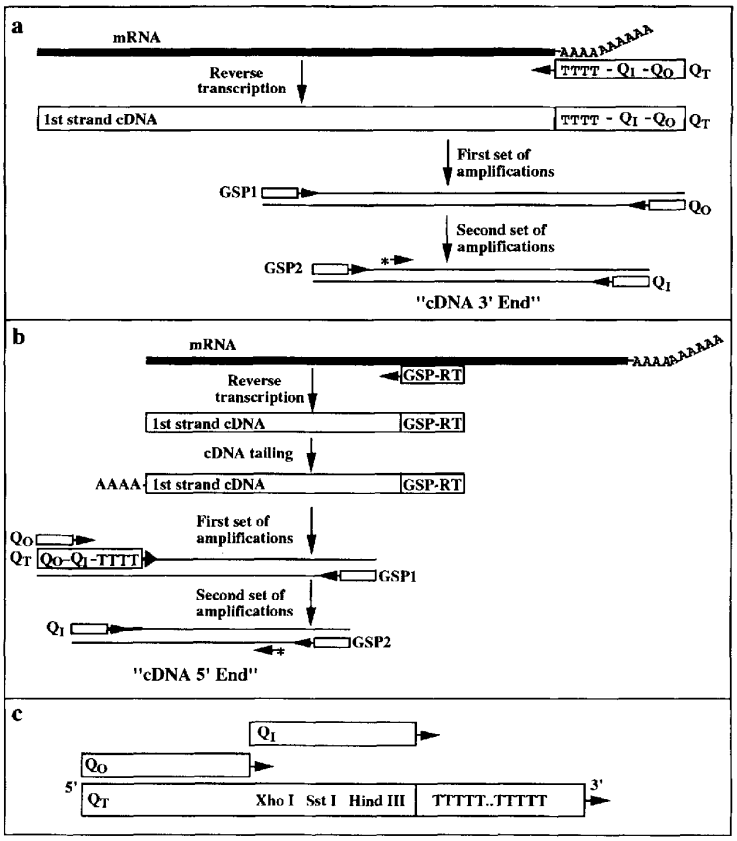


FIGURE 1 Schematic representation of classic RACE. Explanations are given in the text. At each step, the diagram is simplified to illustrate only how the new product formed during the previous step is utilized. (GSP1) Gene-specific primer 1; (GSP2) gene-specific primer 2; (GSP–RT) gene-specific primer used for reverse transcription; (*→)GSP–Hyb/Seq or gene-specific primer for use in hybridization and sequencing reactions. (*a*) Amplification of 3′ partial cDNA ends; (*b*) amplification of 5′ partial cDNA ends; (*c*) schematic representation of the primers used in classic RACE. The 52-nucleotide Q_T primer (5′ Q_O–Q_I–TTTT 3′) contains a 17-nucleotide oligo(dT) sequence at the 3′ end followed by a 35-nucleotide sequence encoding *Hin*dIII, *Sst*I, and *Xho*I recognition sites. The Q_I and Q_O primers overlap by 1 nucleotide; the Q_I primer contains all three of the recognition sites.

Primers:
Q_T: 5′-CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGCTTTTTTTTTTTTTTTT-3′
Q_O:5′-CCAGTGAGCAGAGTGACG-3′
Q_I: 5′-GAGGACTCGAGCTCAAGC-3′

To generate "3′ end" partial cDNA clones, mRNA is reverse transcribed using a "hybrid" primer (Q_T) that consists of 17 nucleotides of oligo(dT) followed by a unique 35-base oligonucleotide sequence (Q_I–Q_O; Fig. 1a and c), which in many reports is denoted as an "anchor" primer. Amplification is then performed using a primer containing part of this sequence (Q_O) that now binds to each cDNA at its 3′ end, and using a primer derived from the gene of interest (GSP1). A second set of amplification cycles is then carried out using "nested" primers (Q_I and GSP2) to quench the amplification of nonspecific products. To generate "5′-end" partial cDNA clones, reverse transcription (primer extension) is carried out using a gene-specific primer (GSP–RT; Fig. 1b) to generate first-strand products. Then, a poly(A) tail is appended

Seikagaku Corporation Exhibit 2004 - Page 84 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

using terminal deoxynucleotidyltransferase (TdT) and dATP. Amplification is then achieved using (1) the hybrid primer Q_T to form the second strand of cDNA, (2) the Q_O primer, and (3) a gene-specific primer upstream of the one used for reverse transcription. Finally, a second set of PCR cycles is carried out using nested primers (Q_I and GSP2) to increase specificity.[5]

Classic RACE Variations

In general, as described above, the gene-specific primer is derived from a short stretch of sequence that is already known from the mRNA of interest. A frequent question is whether degenerate primers, that is, those directed against predicted nucleotide sequence based on known amino acid sequence, can be used instead. Although such primers increase the quantity of spurious amplification, the approach can work, if other parameters are favorable (i.e., message abundance, GC composition, and cDNA end size; see Monstein et al.).[14]

At the unknown end of the cDNA, the 5′ end can be tailed with C's instead of A's, and then amplified using a hybrid primer with a tail containing G's,[2] or a mixture of G's and inosines (I).[11] Although the G:I approach entails synthesizing a primer that can be used for 5′ RACE only [because a T-tailed primer must be used to anneal to the poly(A) tail of the 3′ end], there may be sufficient benefits from using a mixed G:I tail to justify the cost, as the G:I region should anneal at temperatures similar to those of other primers normally used in PCR. In contrast, it is believed that homopolymers of either T's or G's present problems during PCR, because of the very low and very high annealing temperatures, respectively, required for their optimal usage.[1,11] On the other hand, the inosine residues function as degenerate nucleotides and will lead to higher spurious amplification, so the magnitude of the benefit of using a mixed G:I primer is unknown.

To minimize the length of homopolymer tail actually amplified, a "Lock-Docking" primer was developed by Borson et al.[8] In this approach, the final 2 nucleotides on the 3′ end of the primer are degenerate. For example, to amplify cDNAs linked to an A-tail, the lock-docking primer would look like

$$5'\text{-XXXXXXXXXX-AAAAAAAAA-}\begin{pmatrix}G\\T\\C\end{pmatrix}\begin{pmatrix}G\\A\\T\\C\end{pmatrix}\text{-}3'$$

where X represents (e.g.) one or more restriction sites at the 5′ end of the primer. The advantage of this approach is that it forces the primer to anneal to the junction of the natural or appended homopolymer tail and the cDNA sequence. The disadvantage is that it is necessary to synthesize four primers, as most synthesizers can only synthesize primers starting from an unambiguous 3′ end.

In another variation, the location of the anchor primer is changed from the end of the unknown region of sequence to random points within the unknown region.[7] This is accomplished using a primer containing an anchor region followed by six random nucleotides (5′-XXXXXXX-NNNNNN-3′) either for reverse transcription (3′ RACE) or for creation of the second strand of cDNA (5′ RACE). This approach is valuable when the 3′ or 5′ ends lie so far away from the region of known sequence that the entire unknown region cannot be amplified effectively. Using this approach, cDNA ends of defined sizes are not generated; instead, a library of randomly sized fragments is

Seikagaku Corporation Exhibit 2004 - Page 85 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

obtained, all of which initiate at the gene-specific primer. The largest fragments can be cloned and characterized, extending the length of the known sequence, and the process (or standard RACE) repeated until the real unknown end is identified. The development of "long" PCR may make this approach unnecessary.

New RACE

The most technically challenging step in classic 5′ RACE is to cajole reverse transcriptase to copy the mRNA of interest in its entirety into first-strand cDNA. Because prematurely terminated first-strand cDNAs are tailed by terminal transferase just as effectively as full-length cDNAs, cDNA populations composed largely of prematurely terminated first strands will result primarily in the amplification and recovery of cDNA ends that are not full length either (Fig. 2a). This problem is encountered routinely for vertebrate genes, which are often quite GC rich at their 5′ ends and thus frequently contain sequences that hinder reverse transcription. A number of laboratories have developed steps or protocols designed to approach the problem,[17–24] the protocol described here and denoted "new RACE" is for the most part a composite adapted from the cited reports.

New RACE departs from classic RACE in that the "anchor" primer is attached to the 5′ end of the mRNA **before** the reverse transcription step. Hence, the anchor sequence becomes incorporated into the first-strand cDNA if the reverse transcription proceeds through the entire length of the mRNA of interest (and through the relatively short anchor sequence) as shown in Figure 2b.

Before beginning new RACE (Fig. 3a), the mRNA is subjected to a dephos-

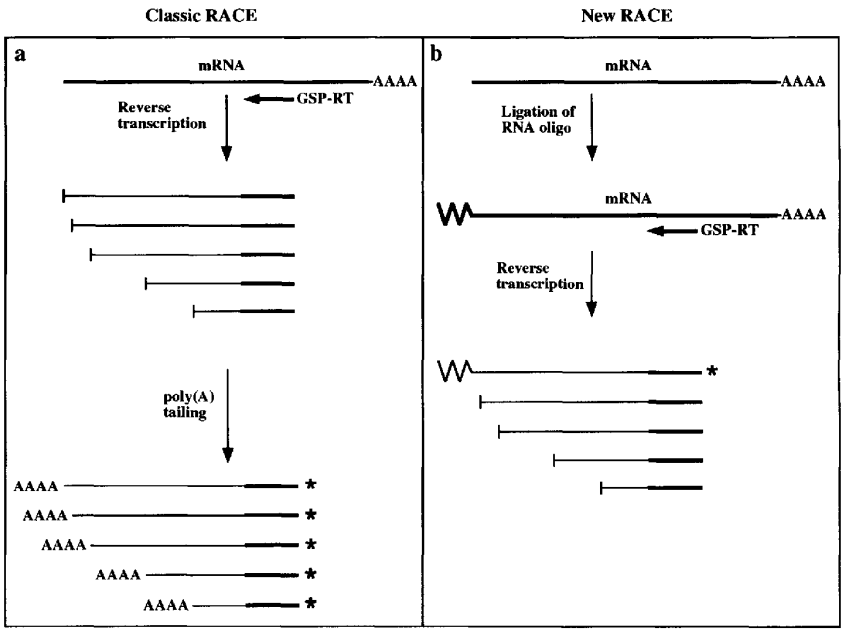


FIGURE 2 A depiction of the advantage of using new RACE over classic RACE. (*a*) In classic RACE, premature termination in the reverse transcription step results in polyadenylation of less than full-length first-strand cDNAs, all of which can be amplified using PCR to generate less than full-length cDNA 5′ ends. (*) cDNA ends created that will be amplified in the subsequent PCR reaction. (*b*) In new RACE, less than full-length cDNAs are also created but are not terminated by the anchor sequence; hence, they cannot be amplified in the subsequent PCR reaction.

Seikagaku Corporation Exhibit 2004 - Page 86 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

phorylation step using calf intestinal phosphatase (CIP). This step actually does nothing to full-length mRNAs, which have methyl-G caps at their termini; but it does dephosphorylate degraded mRNAs, which are uncapped at their termini.[19] This makes the degraded RNA biologically inert during the ensuing ligation step, because the phosphate group is required to drive the reaction. The full-length mRNAs are then decapped using tobacco acid pyrophosphatase (TAP), which leaves them with an active and phosphorylated 5′ terminus.[18,22] Using T4 RNA ligase, this mRNA is then ligated to a short synthetic RNA oligonucleotide that has been generated by in vitro transcription of a linearized plasmid (Fig. 3b).[17] The RNA oligonucleotide mRNA hybrids are then reverse transcribed using a gene-specific primer or random primers to create first-strand cDNA. Finally, the 5′ cDNA ends are amplified in two nested PCR reactions using additional gene-specific primers and primers derived from the sequence of the RNA oligonucleotide.

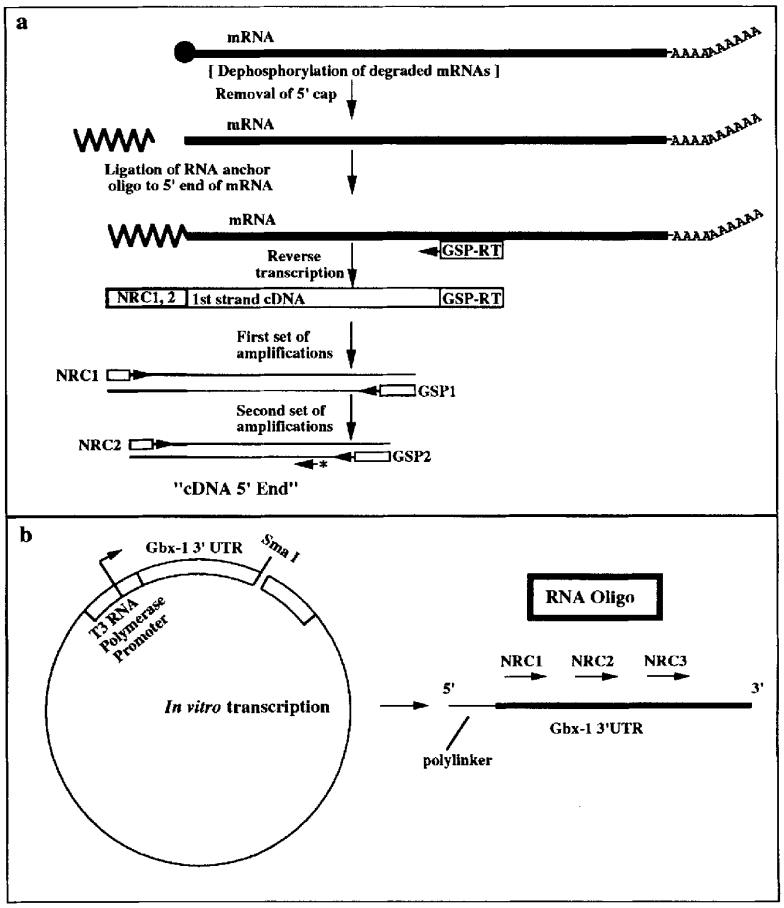


FIGURE 3 Schematic representation of new RACE. Explanations are given in the text. At each step, the diagram is simplified to illustrate only how the new product formed during the previous step is utilized. See the legend to Fig. 1 for description of some primers. (*a*) Amplification of 5′ partial cDNA ends; (*b*) In vitro synthesis of the RNA oligonucleotide used for ligation in new RACE and schematic representation of the corresponding required primers. A 132-nucleotide RNA oligonucleotide is produced by in vitro transcription of the plasmid depicted using T3 RNA polymerase. Primers NRC-1, NRC-2, and NRC-3 are derived from the sequence of the oligonucleotide but do not encode restriction sites. To assist in the cloning of cDNA ends, the sequence "ATCG" is added to the 5′ end of NRC-2, as described in the cloning section of the text.

Seikagaku Corporation Exhibit 2004 - Page 87 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

The new RACE approach can also be used to generate 3′ cDNA ends[19] (see related protocols),[18,20] and is useful in particular for nonpolyadenylated RNAs. In brief, cytoplasmic RNA is dephosphorylated and ligated to a short synthetic RNA oligonucleotide as described above. Although ligation of the oligonucleotide to the 5′ end of the RNA was emphasized above, RNA oligonucleotides actually ligate to both ends of the cytoplasmic RNAs. For the reverse transcription step, a primer derived from the RNA oligonucleotide sequence is used (e.g., the reverse complement of NRC-3, Fig. 3). Reverse transcription of the RNA oligonucleotides that happen to be ligated to the 3′ end of the cytoplasmic RNAs results in the creation of cDNAs that have the RNA oligonucleotide sequence appended to their 3′ end. Gene-specific primers oriented in the 5′ → 3′ direction and new RACE primers (e.g., the reverse complements of NRC-2 and NRC-1; Fig. 3) can be used in nested PCR reactions to amplify the 3′ ends.

Protocols

CLASSIC RACE

Materials

The materials required for this procedure can be purchased, along with the appropriate 5× or 10× enzyme reaction buffers, from most major suppliers. MMLV reverse transcriptase can be obtained from Bethesda Research Labs (BRL), RNasin and *Taq* polymerase from Promega Biotech, and TdT from either BRL or Boehringer Mannheim. Enzymes are used as directed by the suppliers, except for *Taq* polymerase. Instead of using the recommended reaction mixture, a 10× buffer consisting of 670 mM Tris-HCl (pH 9.0), 67 mM $MgCl_2$, 1700 μg/ml of BSA, and 166 mM $(NH_4)_2SO_4$ is substituted, and reaction conditions are altered as described further below.[1,16] Oligonucleotide primer sequences are listed in the legend to Figure 1. Primers can be used "crude" except for Q_T, which should be purified to ensure that it is uniformly full length. dNTPs can be purchased as 100 mM solutions from PL-Biochemicals/Pharmacia or Boehringer Mannheim.

3′ End cDNA amplification

Step 1. Reverse Transcription to Generate cDNA Templates **Procedure:** Assemble reverse transcription components on ice: 4 μl of 5× reverse transcription buffer, 1.3 μl of dNTPs (stock concentration is 15 mM of each dNTP), 0.25 μl (10 units) of RNasin, and 0.5 μl of Q_T primer (100 ng/μl). Heat 1 μg of poly(A)$^+$/ RNA or 5 μg of total RNA in 14 μl of water at 80°C for 3 min and cool rapidly on ice. Add to reverse transcription components. Add 1 μl (200 units) of MMLV reverse transcriptase, and incubate for 1 hr at 37°C, 30 min at 42°C, and 10 min at 50°C. Dilute the reaction mixture to 1 ml with TE [10 mM Tris-HCl (pH 7.5)/1 mM EDTA] and store at 4°C (3′ end cDNA pool).

Comments: Poly(A)$^+$ RNA is used preferentially for reverse transcription to decrease background, but it is unnecessary to prepare it if only total RNA is available. An important factor in the generation of full-length 3′ end partial cDNAs concerns the stringency of the reverse transcription reaction. Reverse transcription reactions were historically carried out at relatively low temperatures (37–42°C) using a vast excess of primer (~½ the mass of the mRNA, which represents an ~30:1 molar ratio). Under these low stringency conditions, a stretch of A residues as short as 6–8 nucleotides will suffice as a binding site for an oligo(dT)-tailed primer. This may result in cDNA synthesis

Seikagaku Corporation Exhibit 2004 - Page 88 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

being initiated at sites upstream of the poly(A) tail, leading to truncation of the desired amplification product. It should be assumed that this has occurred if a canonical polyadenylation signal sequence[25] is not found near the 3′ end of the cDNAs generated. This can be minimized by controlling two parameters: primer concentration and reaction temperature. The primer concentration can be reduced dramatically without decreasing significantly the amount of cDNA synthesized[26] and will begin to bind preferentially to the longest A-rich stretches present [i.e., the poly(A) tail]. The quantity recommended above represents a good starting point; it can be reduced fivefold further if significant truncation is observed.

In the protocol described above, the incubation temperature is raised slowly to encourage reverse transcription to proceed through regions of difficult secondary structure. Because the half-life of MMLV rapidly decreases as the incubation temperature increases, the reaction cannot be carried out at elevated temperatures in its entirety. Alternatively, the problem of difficult secondary structure (and nonspecific reverse transcription) can be approached using heat stable reverse transcriptases, which are now available from several suppliers (Perkin-Elmer Cetus, Amersham, Epicentre Technologies, and others). As in PCR reactions, the stringency of reverse transcription can thus be controlled by adjusting the temperature at which the primer is annealed to the mRNA. The optimal temperature depends on the specific reaction buffer and reverse transcriptase used and should be determined empirically but will usually be found to be in the range of 48–56°C for a primer terminated by a 17-nucleotide oligo(dT) tail.

Step 2. Amplification **Procedure:** First round: Add an aliquot of the cDNA pool (1 μl) and primers (25 pmoles each of GSP1 and Q_O) to 50 μl of PCR cocktail (1× *Taq* polymerase buffer, as described above, each dNTP at 1.5 mM and 10% DMSO) in a 0.5-ml microcentrifuge tube, and heat in a DNA thermal cycler for 5 min at 98°C to denature the first-strand products. Cool to 75°C. Add 2.5 units of *Taq* polymerase, overlay the mixture with 30 μl of mineral oil (Sigma 400-5; preheat the oil in the thermal cycler to 75°C, and incubate at the appropriate annealing temperature (52–60°C) for 2 min. Extend the cDNAs at 72°C for 40 min. Carry out 30 cycles of amplification using a step program (94°C, 1 min; 52–60°C, 1 min; 72°C, 3 min), followed by a 15-min final extension at 72°C. Cool to room temperature.

Second round: Dilute a portion of the amplification products from the first round 1:20 in TE. Amplify 1 μl of the diluted material with primers GSP2 and Q_I using the procedure described above, but eliminate the initial 2-min annealing step and the 72°C, 40-min extension step.

Comments: It is important to add the *Taq* polymerase *after* heating the mixture to a temperature above the T_m off the primers (hot start PCR). Addition of the enzyme prior to this point allows one cycle to take place at room temperature, promoting the synthesis of nonspecific background products dependent on low-stringency interactions.

An annealing temperature close to the effective T_m of the primers should be used. The Q_I and Q_O primers work well at 60°C under the PCR conditions recommended here, although the actual optimal temperature may depend on the PCR machine used. Gene-specific primers of similar length and GC content should be chosen. Computer programs to assist in the selection of primers are widely available and should be used. An extension time of 1 min/kb expected product should be allowed during the amplification cycles. If the expected length of product is unknown, try 3–4 min initially.

Very little substrate is required for the PCR reaction: 1 μg of poly(A)$^+$ RNA typically contains ~5×10^7 copies of each low abundance transcript. The PCR reaction described here works optimally when 10^3–10^5 templates (of the de-

Seikagaku Corporation Exhibit 2004 - Page 89 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

sired cDNA) are present in the starting mixture; thus, as little as 0.002% of the reverse transcription mixture suffices for the PCR reaction. Addition of too much starting material to the amplification reaction will lead to production of large amounts of nonspecific product and should be avoided. The RACE technique is particularly sensitive to this problem, as every DNA in the mixture, desired and undesired, contains a binding site for the Q_O and Q_I primers.

It was found empirically that allowing extra extension time (40 min) during the first amplification round (when the second strand of cDNA is created) sometimes resulted in increased yields of the specific product relative to background amplification, and in particular, increased the yields of long cDNAs versus short cDNAs when specific cDNA ends of multiple lengths were present.[1] Prior treatment of cDNA templates with RNA hydrolysis or a combination of RNase H and RNase A infrequently improves the efficiency of amplification of specific cDNAs.

Choice of buffer systems: For some applications intended for cloned PCR products, such as expressing cDNAs to generate proteins, it is critically important to minimize the rate at which mutations occur during amplification. In other applications, such as using the cloned DNA as a probe in hybridization experiments, the presence of a few mutations is relatively unimportant and thus it is most convenient to use PCR conditions that maximize the likelihood of generating the desired product the first time a set of primers is used. Unfortunately, PCR conditions that result in a minimum of mutations can result in erratic yields and often the desired product cannot be generated until the PCR conditions have been optimized, whereas PCR conditions that reliably produce desired products result in a relatively high mutation rate (~1% after 30 rounds). Thus, appropriate conditions must be chosen to generate the PCR products required prior to undertaking cloning steps. PCR conditions that result in a minimum of mutations require the use of nucleotides (dNTPs) at low concentrations (0.2 mM). Using the conditions recommended for *Taq* polymerase by Perkin-Elmer Cetus results in an error rate of ~0.05% after 30 rounds of amplification. However, the conditions recommended often have to be optimized, meaning that the pH of the buffer and the concentration of magnesium have to be adjusted until the desired product is observed. In addition, inclusion of DMSO or formamide may be required. For those who do not wish to prepare their own reagents to carry out optimization experiments, such kits are commercially available (e.g., from Invitrogen and Stratagene). PCR conditions that work much more frequently in the absence of optimization steps require the use of DMSO, ammonium sulfate, and relatively high concentrations (1.5 mM) of dNTPs, as described above. It should be noted that the inclusion of DMSO to 10% decreases primer melting temperatures (and thus optimal annealing temperatures) by ~5–6°C.

5′ End cDNA Amplification

Step 1. Reverse Transcription to Generate cDNA Templates **Procedure:** Assemble reverse transcription components on ice: 4 μl of 5× reverse transcription buffer, 1.3 μl of dNTPs (stock concentration is 15 mM of each dNTP), 0.25 μl (10 units) of RNasin, and 0.5 μl of GSP–RT primer (100 ng/μl). Heat 1 μg of poly(A)$^+$ RNA or 5 μg of total RNA in 14 μl of water at 80°C for 3 min and cool rapidly on ice. Add to reverse transcription components. Add 1 μl (200 units) of MMLV reverse transcriptase and incubate for 1 hr at 37°C, 30 min at 42°C, and 10 min at 50°C. Dilute the reaction mixture to 400 μl with TE [10 mM Tris-HCl (pH 7.5)/1 mM EDTA] and store at 4°C (5′-end nontailed cDNA pool).

Comments: Many of the remarks made above in the section on reverse transcribing 3′ end partial cDNAs are also relevant here and should be noted.

Seikagaku Corporation Exhibit 2004 - Page 90 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

There is, however, one major difference. The efficiency of cDNA extension is now critically important, as each specific cDNA, no matter how short, is subsequently tailed and becomes a suitable template for amplification (Fig. 2a). Thus, the PCR products eventually generated directly reflect the quality of the reverse transcription reaction. Extension can be maximized by using clean, intact RNA, by selecting the primer for reverse transcription to be near the 5′ end of region of known sequence, and in theory by using heat-stable reverse transcriptase at elevated temperatures or a combination of MMLV and heat-stable reverse transcriptase at multiple temperatures. Synthesis of cDNAs at elevated temperatures should diminish the amount of secondary structure encountered in GC-rich regions of the mRNA.

Step 2. Appending a Poly(A) Tail to First-Strand cDNA Products **Procedure:** Remove excess primer using Microcon-100 spin filters (Amicon Corp.) or an equivalent product, following the manufacturer's instructions. Wash the material by spin filtration twice more using TE. The final volume recovered should not exceed 10 μl.

Add 4 μl of 5× tailing buffer [125 mM Tris-HCl (pH 6.6), 1 M KCacodylate, and 1250 μg/ml BSA], 1.2 μl of 25 mM $CoCl_2$, 4 μl of 1 mM dATP, and 10 units of TdT. Incubate for 5 min at 37°C and then 5 min at 65°C. Dilute to 500 μl with TE (5′-end-tailed cDNA pool).

Comments: To attach a known sequence to the 5′ end of the first strand cDNA, a homopolymeric tail is appended using TdT. We prefer appending poly(A) tails rather than poly(C) tails for several reasons. First, the 3′-end strategy is based on the naturally occurring poly(A) tail; thus, the same adapter primers can be used for both ends, decreasing variability in the protocol and cost. Second, because A/T binding is weaker than G/C binding, longer stretches of A residues (~2×) are required before the oligo(dT)-tailed Q_T primer will bind to an internal site and truncate the amplification product. Third, vertebrate coding sequences and 5′ untranslated regions tend to be biased toward G/C residues; thus, use of a poly(A) tail further decreases the likelihood of inappropriate truncation.

Unlike many other situations in which homopolymeric tails are appended, the actual length of the tail added here is unimportant, as long as it exceeds 17 nucleotides. This is because although the oligo(dT)-tailed primers subsequently bind all along the length of the appended poly(A) tail, only the innermost one becomes incorporated into the amplification product, and consequently, the remainder of the poly(A) tail is lost.[1] The truncation appears to happen because *Taq* polymerase is unable to resolve branched structures (efficiently). The conditions described in the procedure above result in the addition of 30–400 nucleotides.

Step 3. Amplification **Procedure:** First round: Add an aliquot of the 5′ end tailed cDNA pool (1 μl) and primers (25 pmoles each of GSP1 and Q_O, as shown in Fig. 2b; and 2 pmoles of Q_T) to 50 μl of PCR cocktail (1× *Taq* polymerase buffer, each dNTP at 1.5 mM, and 10% DMSO) in a 0.5-ml microcentrifuge tube, and heat in a DNA thermal cycler for 5 min at 98°C to denature the first-strand products. Cool to 75°C. Add 2.5 units of *Taq* polymerase, overlay the mixture with 30 μl of mineral oil (Sigma 400-5; preheat it in the thermal cycler to 75°C), and incubate at the appropriate annealing temperature (48–52°C) for 2 min. Extend the cDNAs at 72°C for 40 min. Carry out 30 cycles of amplification using a step program (94°C, 1 min; 52–60°C, 1 min; 72°C, 3 min), followed by a 15-min final extension at 72°C. Cool to room temperature.

Second round: Dilute a portion of the amplification products from the first-round 1:20 in TE. Amplify 1 μl of the diluted material with primers GSP2

Seikagaku Corporation Exhibit 2004 - Page 91 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

and Q_I using the procedure described above, but eliminate the initial 2-min annealing step and the 72°C, 40-min extension step.

Comments: Many of the remarks made above in the section on amplifying 3′ end partial cDNAs are also relevant here and should be noted. There is, however, one major difference. The annealing temperature in the first step (48–52°C) is lower than that used in successive cycles (52–60°C). This is because cDNA synthesis during the first round depends on the interaction of the appended poly(A) tail and the oligo(dT)-tailed Q_T primer, whereas in all subsequent rounds, amplification can proceed using the Q_O primer, which is composed of ~60% GC and can anneal at a much higher temperature to its complementary target.

NEW RACE

Materials

See the Classic RACE (above) for sources of some of the required materials. TAP and a cost-effective RNA transcription kit can be purchased along with the appropriate 10× enzyme reaction buffers from Epicentre (Madison, WI). T4 RNA ligase can be purchased from New England Biolabs or Boehringer Mannheim. Note that 10× T4 RNA ligase buffers supplied by some manufacturers contain too much ATP.[17] Check the composition of any commercially supplied 10× buffer and make your own if it contains more than 1 mM ATP (final 1× concentration should be 0.1 mM), as described below.

5′ End cDNA amplification

The procedure detailed below is described using relatively large amounts of RNA and can be scaled down if RNA quantities are limited. The advantage of starting with large amounts of RNA is that aliquots can be electrophoresed quickly after each step of the procedure to confirm that detectable degradation of the RNA has not occurred, and the dephosphorylated, decapped, ligated RNA can be stored indefinitely for many future experiments.

Step 1. Dephosphorylation of degraded RNAs **Procedure:** In general, follow the manufacturer's recommendations for use of the phosphatase. To dephosphorylate RNA using CIP from Boehringer Mannheim, prepare a reaction mixture containing (50 μg of RNA in 41 μl of water, 5 μl of 10× buffer, 0.5 μl of 100 mM DTT, 1.25 μl of RNasin (40 U/μl), and 3.5 μl of CIP (1 U/μl). Incubate the reaction at 50°C for 1 hr, and add proteinase K (Boehringer Mannheim) to 50 μg/ml and incubate at 37°C for 30 min. Extract the reaction with a mixture of phenol/chloroform, extract again with chloroform, and precipitate the RNA using 1/10 volume of 3 M NaOAc and 2.5 volumes of ethanol. Resuspend the RNA in 43.6 μl of water. Electrophorese 2 μg (1.6 μl) on a 1% TAE agarose gel adjacent to a lane containing 2μg of the original RNA preparation, stain the gel with ethidium bromide, and confirm visually that the RNA remained intact during the dephosphorylation step.

Step 2. Decapping of intact RNAs **Procedure:** Prepare a reaction mixture containing 38 μg of RNA in 42 μl of water (this is the RNA recovered from step 1), 5 μl of 10× TAP buffer; 1.25 μl of RNasin, (40 U/2μl), 1μl of 100 mM ATP; and 1 μl of TAP (5 U/μl). Incubate the reaction at 37°C for 1 hr and add 200 μl of TE. Extract the reaction with a mixture of phenol/chloroform, extract again with chloroform, and precipitate the RNA using 1/10 volume of 3 M NaOAc and 2.5 volumes of ethanol. Resuspend the RNA in 40 μl of water.

Seikagaku Corporation Exhibit 2004 - Page 92 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

Electrophorese 2 μg on a TAE 1% agarose gel adjacent to a lane containing 2 μg of the original RNA preparation, stain the gel with ethidium bromide, and confirm visually that the RNA remained intact during the decapping step. **Comment:** Most protocols call for much more TAP. The excess is not necessary, and the enzyme is very expensive.

Step 3. Preparation of RNA oligonucleotides Choose a plasmid that can be linearized at a site ~100 bp downstream from a T7 or T3 RNA polymerase site (see Fig. 3b). Ideally, a plasmid containing some insert cloned into the first polylinker site is optimal, because primers made from palindromic polylinker DNA do not perform well in PCR. For my experiments, I use the 3′ UTR of the mouse gene Gbx-1,[27] which is cloned into the *Sst*I site of pBS-SK (Stratagene). I linearize with *Sma*I and transcribe with T3 RNA polymerase to produce a 132-nucleotide RNA oligonucleotide, of which all but 17 nucleotides are from Gbx-1. Note that adenosines are the best acceptors for the 3′ end of the RNA oligonucleotide to ligate to the 5′ end of its target, if an appropriate restriction site can be found. The primers subsequently used for amplification are all derived from the Gbx-1 3′ UTR sequence. Interested investigators are welcome to the Gbx-1 NRC primer sequences and plasmid upon request.

Carry out a test transcription to make sure that everything is working; then scale up. The oligonucleotide can be stored at −80°C indefinitely for many future experiments, and it is important to synthesize enough oligonucleotide so that losses due to purification and spot checks along the way will leave plenty of material at the end of the procedure.

Procedure: Linearize 25 μg of the plasmid that is to be transcribed (the plasmid should be reasonably free of RNases). Treat the digestion reaction with 50 μg/ml of Proteinase K for 30 min at 37°C, followed by 2× phenol/$CHCl_3$ extractions, one $CHCl_3$ extraction, and ethanol precipitation. Resuspend the template DNA in 25 μl of TE (pH 8.0) for a final concentration of ~1 μg/μl.

For transcription, mix at room temperature in the following order, and then incubate the reaction mixture at 37°C for 1 hr.

	Test Scale	or	Prep Scale
DEPC–water	4 μl		80
5× buffer	2		40
0.1 M DTT	1		20
10 mM UTP	0.5		10
10 mM ATP	0.5		10
10 mM CTP	0.5		10
10 mM GTP	0.5		10
Restricted DNA (1 μg/μl)	0.5		10
RNasin (40 U/μl)	0.25		5
RNA polymerase (20 U/μl)	0.25		5

Add 0.5 μl DNase (RNase free) for every 20 μl of reaction volume and incubate at 37°C for 10 min. Run 5 μl of test or prep reaction on a 1% TAE agarose gel to check. Expect to see a diffuse band at about the right size (or a bit smaller) in addition to some smearing all up and down the gel.

Purify the oligo by extracting with phenol/$CHCl_3$ and $CHCl_3$, and then clean three times using water and a Microcon spin filter (prerinsed with water). Run another appropriately sized aliquot on a 1% TAE agarose gel to check integrity and concentration of oligo. Microcon 30 spin filters have a cut off size of 60 nucleotides, and Microcon 100 spin filters have a cut off size of 300 nucleotides. Microcon 10 spin filters are probably most appropriate if the oligo is <100 nucleotides, and Microcon 30 spin filters for anything larger.

Seikagaku Corporation Exhibit 2004 - Page 93 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

Step 4. RNA Oligonucletide–cellular RNA Ligation Set up two tubes: one with TAP_{ped} cellular RNA, the other with $unTAP_{ped}$ cellular RNA.

11.25	µl	water
3	µl	10× buffer = 500 mM Tris (pH 7.9), 100 mM $MgCl_2$, 20 mM DTT, 1 mg/ml of BSA
0.75	µl	RNasin (40 U/µl)
2	µl	4 µg of RNA oligonucleotide (3–6 molar excess over target cellular RNA)
10	µl	10 µg of TAP_{ped} (or $unTAP_{ped}$) RNA
1.5	µl	2 mM ATP
1.5	µl	T4 RNA ligase (20 U/µl)
30	µl	

Incubate for 16 hr at 17°C. Purify the ligated oligonucleotide–RNA using Microcon 100 spin filtration (3× in water; prerinse filter with RNase-free water). The volume recovered should not exceed 20 µl. Run one-third of the ligation on a 1% TAE agarose gel to check integrity of the ligated RNA. It should look about as it did before ligation.

Step 5. Reverse Transcription **Procedure:** Assemble reverse transcription components on ice: 4 µl of 5× reverse transcription buffer, 1.3 µl of dNTPs (stock concentration is 15 mM of each dNTP), 0.25 µl (10 units) of RNasin, and 1 µl of antisense-specific primer (20 ng/µl), or random hexamers (100 ng/µl). Heat the remaining RNA (~6.7 µg) in 14 µl of water at 80°C for 3 min, and cool rapidly on ice. Add to reverse transcription components. Add 1 µl (200 units) of MMLV reverse transcriptase and incubate for 1 hr at 37°C, 30 min at 42°C, and 10 min at 50°C. Dilute the reaction mixture to 100 µl with TE [10 mM Tris-HCl (pH 7.5)/1 mM EDTA] and store at 4°C (5′ end oligonucleotide–cDNA pool).

Step 6. Amplification **Procedure:** First round: Add an aliquot of the 5′-end oligonucleotide–cDNA pool (1 µl) and primers (25 pmoles each of GSP1 and NRC-1, as in Fig. 3a) to 50 µl of PCR cocktail (1× *Taq* polymerase buffer, each dNTP at 1.5 mM, and 10% DMSO) in a 0.5-ml microcentrifuge tube, and heat in a DNA thermal cycler for 5 min at 98°C to denature the first-strand products. Cool to 75°C. Add 2.5 units of *Taq* polymerase, overlay the mixture with 30 µl of mineral oil (Sigma 400-5; preheat it in the thermal cycler to 75°C), and incubate at the appropriate annealing temperature (52–60°C) for 2 min. Extend the cDNAs at 72°C for 40 min. Carry out 35 cycles of amplification using a step program (94°C, 1 min; 52–60°C, 1 min; 72°C, 3 min), followed by a 15-min final extension at 72°C. Cool to room temperature.

Second round: Dilute a portion of the amplification products from the first round 1:20 in TE. Amplify 1 µl of the diluted material with primers GSP2 and NRC-2 using the procedure described above, but eliminate the initial 2-min annealing step, and the 72°C, 40-min extension step.

Comments: Many of the remarks made in the Classic RACE (above) on amplifying 5′-end partial cDNAs are also relevant here and should be noted.

Analysis of Amplification Products

The production of specific partial cDNAs by the RACE protocol is assessed using Southern blot hybridization analysis. After the second set of amplification cycles, the first and second set reaction products are electrophoresed in a 1% agarose gel, stained with ethidium bromide (EtBr), denatured, and transferred to a nylon membrane. After hybridization with a labeled oligomer or gene fragment derived from a region contained within the amplified fragment (e.g., GSP-Hyb/Seq in Fig. 1a,b), gene-specific partial cDNA ends should be detected easily. Yields of the desired product relative to nonspecific am-

Seikagaku Corporation Exhibit 2004 - Page 94 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

plified cDNA in the first round products should vary from <1% of the amplified material to nearly 100%, depending largely on the stringency of the amplification reaction, the amplification efficiency of the specific cDNA end, and the relative abundance of the specific transcript within the mRNA source. In the second set of amplification cycles, ~100% of the cDNA detected by EtBr staining should represent specific product. If specific hybridization is not observed, then troubleshooting steps should be initiated.

Information gained from this analysis should be used to optimize the RACE procedure. If low yields of specific product are observed because nonspecific products are being amplified efficiently, then annealing temperatures can be raised gradually (~2°C at a time) and sequentially in each stage of the procedure until nonspecific products are no longer observed. Alternatively, some investigators have reported success using the touchdown PCR procedure to optimize the annealing temperature without trial and error.[28] Optimizing the annealing temperature is also indicated if multiple species of specific products are observed, which could indicate that truncation of specific products is occurring. If multiple species of specific products are observed after the reverse transcription and amplification reactions have been fully optimized, then the possibility should be entertained that alternate splicing or promoter use is occurring.

CLASSIC RACE ONLY

If a nearly continuous smear of specific products is observed up to a specific size limit after 5′ end amplification, this suggests that polymerase pausing occurred during the reverse transcription step. To obtain nearly full-length cDNA ends, the amplification mixture should be electrophoresed and the longest products recovered by gel isolation. An aliquot of this material can then be reamplified for a limited number of cycles.

NEW RACE ONLY

Expect to see one or two extra nucleotides insert between the RNA oligonucleotide 3′ end, and the 5′ end of the gene of interest—these come from the transcription step using T7, T3, or SP6, which can add an extra nucleotide or two to oligonucleotides past the end of the template (template-independent transcription).

Compare the results from $unTAP_{ped}$ RNA versus TAP_{ped} RNA. Junction sites (where the oligonucleotide is connected to the 5′ end of your gene) in common arise from ligation of the oligonucleotide to degraded RNA; unique junctions in the TAP_{ped} RNA population represent candidate transcription start sites. If you have RNA degradation sites (e.g., TTT▼AAA) in your 5′ RNA end, you may have substantial numbers of clones that begin at exactly the same nucleotide but arise from ligation of the oligonucleotide to degraded RNA molecules, not from ligation of the oligonucleotide to the true 5′ end of the RNA.

Look for TATA, CCAAT, and initiator element (Inr) sites at or around your candidate transcription site in the genomic DNA sequence if it is available—you should usually be able to find either a TATA or an Inr.

FURTHER ANALYSIS AND USE OF RACE PRODUCTS

Cloning

RACE products can be cloned like any other PCR products.

Option 1 To clone the cDNA ends directly from the amplification reaction (or after gel purification, which is recommended), ligate an aliquot of the

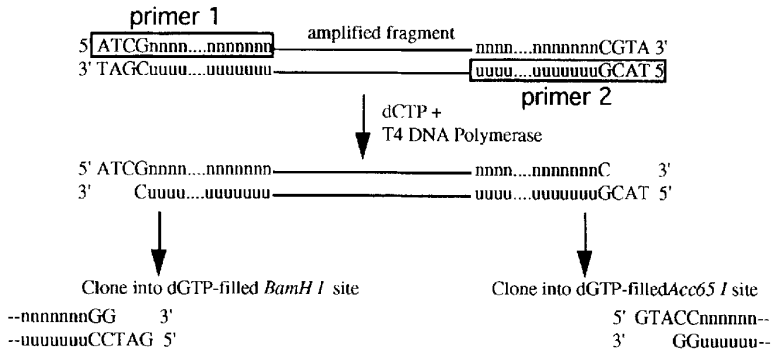
Seikagaku Corporation Exhibit 2004 - Page 95 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

products to plasmid vector encoding a 1-nucleotide 3′ overhang consisting of a T on both strands. Such vector DNA is available commercially (TA Kit, Invitrogen) or can be prepared easily and inexpensively.[16,29–32]

Option 2 The classic RACE Q_I primer encodes *Hin*dIII, *Sst*I, and *Xho*I restriction enzyme sites. Products can be cloned efficiently into vectors that have been double-cut with one of these enzymes and with a blunt-cutting enzyme such as *Sma*I (Note: Remember to "polish" the amplification products with Klenow enzyme or T4 DNA polymerase and separate them from residual *Taq* polymerase and dNTPs before carrying out the restriction enzyme digest). If clones are not obtained, determine whether the restriction enzyme chosen is cutting the amplified gene fragment a second time, at some internal location in the new and unknown sequence. A somewhat easier strategy is to append a restriction site (**not** *Hin*dIII, *Sst*I, or *Xho*I) onto the 5′ end of the GSP2 primer to allow for the creation of overhanging strands at both ends of the amplified product.

Option 3 A safer and very effective approach is to modify the ends of the primers to allow the creation of overhanging ends using T4 DNA polymerase to chew back a few nucleotides from the amplified product in a controlled manner and Klenow enzyme (or Sequenase) to partially fill in restriction enzyme-digested overhanging ends on the vector, as shown below (adapted from Stoker and Iwahana et al.).[33,34] For another conceptual variation, see Rashtchian et al.[10]



Note: *Acc*65I recognizes the same sequence that *Kpn*I does, but creates a 5′ instead of a 3′ overhang.

The advantages of this approach are that (1) it eliminates the possibility that the restriction enzymes chosen for the cloning step will cleave the cDNA end in the unknown region, (2) vector dephosphorylation is not required because vector self-ligation is no longer possible, (3) which means that insert kinasing (and polishing) is not necessary, (4) which means that insert multimerization and fusion clones are not observed either. In addition, the procedure is more reliable than TA cloning.

Procedure

Insert Preparation Select a pair of restriction enzymes for which you can synthesize half-sites appended to PCR primers that can be chewed back to form the appropriate overhangs, as shown for *Bam*HI and *Acc*65I, above. *Not*1 and *Eag*I also are compatible with *Bam*HI and *Acc*65I. For RACE cloning, add

Seikagaku Corporation Exhibit 2004 - Page 96 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

ATCG to the 5′ end of Q_I or NRC-2, and add TACG to the 5′ end of GSP2. Carry out PCR as usual.

After PCR, add proteinase K (10 mg/ml stock) to the PCR reaction to a concentration of 50 μg/ml and incubate at 37°C for 30 min, to remove sticky *Taq* polymerase from the amplified DNA.[35] Extract the PCR products with phenol/$CHCl_3$, then $CHCl_3$ (but do not precipitate), to remove proteins. Filter the PCR products through a Microcon 100 spin column (or Microcon 30 if your product is <150 bp) three times using TE (not water) as the wash buffer, to remove unwanted organics, primers, and dNTPs.

Then, on ice, add the selected dNTP(s) (e.g., dCTP) to a final concentration of 0.2 mM, 1/10 volume of 10× T4 DNA polymerase buffer, and 1–2 units of T4 DNA polymerase. Incubate at 12°C for 15 min, and then at 75°C for 10 min to heat inactivate the T4 DNA polymerase. (Optional: Gel isolate DNA fragment of interest, depending on degree of success of PCR amplification.)

Vector Preparation Digest vector (e.g., pBluescript, Stratagene) using the selected enzymes (e.g., *Acc*65I and *Bam*HI) under optimal conditions, in a volume of 10 μl. Then, add a 10-μl mixture containing the selected dNTP(s) (e.g., dGTP) at a final concentration of 0.4 mM, 1 μl of the restriction buffer used for digestion, 0.5 μl of Klenow, and 0.25 μl of Sequenase. Incubate at 37°C for 15 min, and then 75°C for 10 min to heat inactivate the polymerases. Gel-isolate the linearized vector fragment. For ligation, use equal molar amounts of vector and insert.

Sequencing

RACE products can be sequenced directly on a population level using a variety of protocols, including cycle sequencing, from the end at which the gene-specific primers are located. Note that the classic RACE products can not be sequenced on a population level using the Q_I primer at the unknown end, because individual cDNAs contain different numbers of A residues in their poly(A) tails and as a consequence, the sequencing ladder falls out of register after reading through the tail. The 3′-end products can be sequenced from their unknown end using the following set of primers (TTTTTTTTTTTTTTTTA, TTTTTTTTTTTTTTTTG, TTTTTTTTTTTTTTTTC). The non-T nucleotide at the 3′ end of the primer forces the appropriate primer to bind to the inner end of the poly(A) tail.[36] The other two primers do not participate in the sequencing reaction. Individual cDNA ends, once cloned into a plasmid vector, can be sequenced from either end using gene-specific or vector primers.

Hybridization Probes RACE products are generally pure enough so that they can be used as probes for RNA and DNA blot analyses. It should be kept in mind that small amounts of contaminating nonspecific cDNAs will always be present. It is also possible to include a T7 RNA polymerase promoter in one or both primer sequences and to use the RACE products in in vitro transcription reactions to produce RNA probes.[5] Primers encoding the T7 RNA polymerase promoter sequence do not appear to function as amplification primers as efficiently as those listed in the legend to Figure 1 (M. Frohman, pers. observation). Thus, the T7 RNA polymerase promoter sequence should not be incorporated into RACE primers as a general rule.

Construction of Full-length cDNAs It is possible to use the RACE protocol to create overlapping 5′ and 3′ cDNA ends that can later, through judicious choice of restriction enzyme sites, be joined together through subcloning to form a full-length cDNA. It is also possible to use the sequence information gained from acquisition of the 5′ and 3′ cDNA ends to make new primers representing the extreme 5′ and 3′ ends of the cDNA, and to employ them to

Seikagaku Corporation Exhibit 2004 - Page 97 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

amplify a de novo copy of a full-length cDNA directly from the 3′-end cDNA pool. Despite the added expense of making two more primers, there are several reasons why the second approach is preferred.

First, a relatively high error rate is associated with the PCR conditions for which efficient RACE amplification takes place, and numerous clones may have to be sequenced to identify one without mutations. In contrast, two specific primers from the extreme ends of the cDNA can be used under inefficient but low-error rate conditions[37] for a minimum of cycles to amplify a new cDNA that is likely to be free of mutations. Second, convenient restriction sites are often not available, thus making the subcloning project difficult. Third, by using the second approach, the synthetic poly(A) tail (if present) can be removed from the 5′ end of the cDNA. Homopolymer tails appended to the 5′ ends of cDNAs have, in some cases, been reported to inhibit translation. Finally, if alternate promoters, splicing, and polyadenylation signal sequences are being used and result in multiple 5′ and 3′ ends, it is possible that one might join two cDNA halves that are never actually found together in vivo. Employing primers from the extreme ends of the cDNA as described confirms that the resulting amplified cDNA represents an mRNA actually present in the starting population.

TROUBLESHOOTING AND CONTROLS

Problems With Reverse Transcription and Prior Steps

Damaged RNA Electrophorese RNA in 1% formaldehyde minigel and examine integrity of the 18S and 28S ribosomal bands. Discard the RNA preparation if ribosomal bands are not sharp.

Contaminants Ensure that the RNA preparation is free of agents that inhibit reverse transcription, for example, LiCl and SDS (see Sambrook et al. regarding the optimization of reverse transcription reactions).[38]

Bad Reagents To monitor reverse transcription of the RNA, add 20 μCi of [^{32}P]dCTP to the reaction, separate newly created cDNAs using gel electrophoresis, wrap the gel in Saran Wrap, and expose it to X-ray film. Accurate estimates of cDNA size can be best determined using alkaline agarose gels, but a simple 1% agarose minigel will suffice to confirm that reverse transcription took place and that cDNAs of reasonable length were generated. Note that adding [^{32}P]dCTP to the reverse transcription reaction results in the detection of cDNAs synthesized both through the specific priming of mRNA and through RNA self-priming. When a gene-specific primer is used to prime transcription (5′ end RACE) or when total RNA is used as a template, the majority of the labeled cDNA will actually have been generated from RNA self-priming. To monitor extension of the primer used for reverse transcription, label the primer using T4 DNA kinase and [γ-^{32}P]ATP prior to reverse transcription. Much longer exposure times will be required to detect the labeled primer extension products than when [^{32}P]dCTP is added to the reaction.

To monitor reverse transcription of the gene of interest, one may attempt to amplify an internal fragment of the gene containing a region derived from two or more exons, if sufficient sequence information is available.

Problems With Tailing

Bad Reagents Tail 100 ng of a DNA fragment of ~100–300 bp for 30 min. In addition, mock-tail the same fragment (add everything but the TdT). Run

Seikagaku Corporation Exhibit 2004 - Page 98 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

both samples in a 1% agarose minigel. The mock-tailed fragment should run as a tight band. The tailed fragment should have increased in size by 20–200 bp and should appear to run as a diffuse band that trails off into higher molecular weight products. If this not observed, replace reagents.

Mock-tail 25% of the cDNA pool (add everything but the TdT) Dilute to the same final concentration as the tailed cDNA pool. This serves two purposes. First, while amplification products will be observed using both tailed and untailed cDNA templates, the actual pattern of bands observed should be different. In general, discrete bands are observed using untailed templates after the first set of cycles, and a broad smear of amplified cDNA accompanied by some individual bands is typically observed using tailed templates. If the two samples appear different, this confirms that tailing took place and that the oligo(dT)-tailed Q_T primer is annealing effectively to the tailed cDNA during PCR. Second, observing specific products in the tailed amplification mixture that are not present in the untailed amplification mixture indicates that these products are being synthesized off the end of an A-tailed cDNA template, rather than by annealing of the dT-tailed primer to an A-rich sequence in or near the gene of interest.

Problems with Amplification

No Product If no products are observed for the first set of amplifications after 30 cycles, add fresh *Taq* polymerase and carry out an additional 15 rounds of amplification (extra enzyme is not necessary if the entire set of 45 cycles is carried out without interruption at cycle 30). Product is always observed after a total of 45 cycles if efficient amplification is taking place. If no product is observed, carry out a PCR reaction using control templates and primers to ensure the integrity of the reagents.

Smeared Product From the Bottom of the Gel to the Loading Well Too many cycles, or too much starting material.

NonSpecific Amplification, But No Specific Amplification Check sequence of cDNA and primers. If all are correct, examine primers (using computer program) for secondary structure and self-annealing problems. Consider ordering new primers. Determine whether too much template is being added, or if the choice of annealing temperatures could be improved.

Alternatively, secondary structure in the template may be blocking amplification. Consider adding formamide[39] or ^{7}aza-GTP (in a 1:3 ratio with dGTP) to the reaction to assist polymerization. ^{7}aza-GTP can also be added to the reverse transcription reaction.

The Last Few Base Pairs of the 5′-End Sequence do not Match the Corresponding Genomic Sequence Be aware that reverse transcriptase and T7 and T3 RNA Polymerase can add on a few extra template-independent nucleotides.

Inappropriate Templates To determine whether the amplification products observed are being generated from cDNA or whether they derive from residual genomic DNA or contaminating plasmids, pretreat an aliquot of the RNA with RNase A.

CONCLUSIONS

The RACE protocol offers several advantages over conventional library screening to obtain additional sequence for cDNAs already partially cloned.

Seikagaku Corporation Exhibit 2004 - Page 99 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

RACE is cheaper, much faster, requires very small amounts of primary material, and provides rapid feedback on the generation of the desired product. Information regarding alternate promoters, splicing, and polyadenylation signal sequences can be obtained and a judicious choice of primers (e.g., within an alternately spliced exon) can be used to amplify a subpopulation of cDNAs from a gene for which the transcription pattern is complex. Furthermore, differentially spiced or initiated transcripts can be separated by electrophoresis and cloned separately, and essentially unlimited numbers of independent clones can be generated to examine rare events. Finally, for 5′ end amplification, the ability of reverse transcriptase to extend cDNAs all the way to the ends of the mRNAs is greatly increased, as a primer extension library is created instead of a general-purpose one.

ACKNOWLEDGMENTS

The work presented here was supported by a grant from the National Institutes of Health (HD29758).

REFERENCES

1. Frohman, M.A., M.K. Dush, and G.R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* **85:** 8998–9002.
2. Loh, E.L., J.F. Elliott, S. Cwirla, L.L. Lanier, and M.M. Davis. 1989. Polymerase chain reaction with single sided specificity: Analysis of T cell receptor delta chain. *Science* **243:** 217–220.
3. Ohara, O., R.I. Dorit, and W. Gilbert. 1989. One-sided PCR: The amplification of cDNA. *Proc. Natl. Acad. Sci.* **86:** 5673–5677.
4. Frohman, M.A. 1989. Creating full-length cDNAs from small fragments of genes: Amplification of rare transcripts using a single gene-specific oligonucleotide primer. In *PCR protocols and applications: A laboratory manual* (ed. M. Innis, D. Gelfand, J. Sninsky, and .. White), pp. 28–38. Academic Press, New York.
5. Frohman, M.A. and G.R. Martin. 1989. Rapid amplification of cDNA ends using nested primers. *Techniques* **1:** 165–173.
6. Dumas, J.B., M. Edwards, J. Delort, and J. Mallet. 1991. Oligodeoxyribonucleotide ligation to single-stranded cDNAs: A new tool for cloning 5′ ends of mRNAs and for constructing cDNA libraries by *in vitro* amplification. *Nucleic Acids Res.* **19:** 5227–5223.
7. Fritz, J.D., M.L. Greaser, and J.A. Wolff. 1991. A novel 3′ extension technique using random primers in RNA-PCR. *Nucleic Acids Res.* **119:** 3747.
8. Borson, N.D., W.L. Salo, and L.R. Drewes. 1992. A lock-docking oligo(dT) primer for 5′ and 3′ RACE PCR. *PCR Methods Applic.* **2:** 144–148.
9. Jain, R., R.H. Gomer, and J.J.J. Murtagh. 1992. Increasing specificity from the PCR-RACE technique. *BioTechniques* **12:** 58–59.
10. Rashtchian, A., G.W. Buchman, D.M. Schuster, and M.S. Berninger. 1992. Uracil DNA glycosylase-mediated cloning of PCR-amplified DNA: Application to genomic and cDNA cloning. *Anal. Biochem.* **206:** 91–97.
11. Schuster, D.M., G.W. Buchman, and A. Rastchian. 1992. A simple and efficient method for amplification of cDNA ends using 5′ RACE. *Focus* **14:** 46–52.
12. Bertling, W.M., F. Beier, and E. Reichenberger. 1993. Determination of 5′ ends of specific mRNAs by DNA ligase-dependent amplification. *PCR Methods Applic.* **3:** 95–99.
13. Frohman, M.A. 1993. Rapid amplification of cDNA for generation of full-length cDNA ends: Thermal RACE. *Methods Enzymol.* **218:** 340–356.
14. Monstein, H.J., J.U. Thorup, R. Folkesson, A.H. Johnsen, and J.F. Rehfeld. 1993. cDNA deduced procionin—structure and expression in protochordates resemble that of procholecystokinin in mammals. *FEBS Lett.* **331:** 60–64.
15. Templeton, N.S., E. Urcelay, and B. Safer. 1993. Reducing artifact and increasing the yield of specific DNA target fragments during PCR-RACE or anchor PCR. *BioTechniques* **15:** 48–50.
16. Frohman, M.A. 1994. Cloning PCR products: Strategies and tactics. In *PCR: The polymerase chain reaction. Methods in molecular biology series* (ed. K.B. Mullis, F. Ferre, R.A. Gibbs), pp. 14–37. Birkhäuser, Boston, MA.
17. Tessier, D.C., R. Brousseau, and T. Vernet. 1986. Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. *Anal. Biochem.* **158:** 171–178.
18. Mandl, C.W., F.X. Heinz, E. Puchhammer-Stockl, and C. Kunz. 1991. Sequencing the termini of capped viral RNA by 5′-3′ ligation and PCR. *BioTechniques* **10:** 484–486.
19. Volloch, V., B. Schweizer, X. Zhang, and S. Rits. 1991. Identification of negative-strand

Seikagaku Corporation Exhibit 2004 - Page 100 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press

complements to cytochrome oxidase subunit III RNA in Trypanosoma brucei. *Biochemistry* **88:** 10671–10675.
20. Brock, K.V., R. Deng, and S.M. Riblet. 1992. Nucleotide sequencing of 5′ and 3′ termini of bovine viral diarrhea virus by RNA ligation and PCR. *Virol. Methods* **38:** 39–46.
21. Bertrand, E., M. Fromont-Racine, R. Pictet, and T. Grange. 1993. Visualization of the interaction of a regulatory protein with RNA *in vivo*. *Proc. Natl. Acad. Sci.* **90:** 3496–3500.
22. Fromont-Racine, M., E. Bertrand, R. Pictet, and T. Grange. 1993. A highly sensitive method for mapping the 5′ termini of mRNAs. *Nucleic Acids Res.* **21:** 1683–1684.
23. Liu, X. and M.A. Gorovsky. 1993. Mapping the 5′ and 3′ ends of tetrahymena-thermophila mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic Acids Res.* **21:** 4954–4960.
24. Sallie, R. 1993. Characterization of the extreme 5′ ends of RNA molecules by RNA ligation-PCR. *PCR Methods Applic.* **3:** 54–56.
25. Wickens, M. and P. Stephenson. 1984. Role of the conserved AAUAAA sequence: Four AAUAAA point mutants prevent mRNA 3′ end formation. *Science* **226:** 1045–1050.
26. Coleclough, C. 1987. Use of primer-restriction end adapters in cDNA cloning. *Methods Enzymol.* **154:** 64–83.
27. Frohman, M.A., M.E. Dickinson, B.L.M. Hogan, and G.R. Martin. 1993. Localization of two new and related homeobox-containing genes to chromosomes 1 and 5, near the phenotypically similar mutant loci *dominant hemimelia* (*Dh*) and *hemimelic extra-toes* (*Hx*). *Mouse Genome* **91:** 323–325.
28. Don, R.H., P.T. Cox, B.J. Wainwright, K. Baker, and J.S. Mattick. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19:** 4008.
29. Mead, D.A., N.K. Pey, C. Herrnstadt, R.A. Marcil, and L.A. Smith. 1991. A universal method for direct cloning of PCR amplified nucleic acid. *Biotechnology* **9:** 657–663.
30. Marchuk, D., M. Drumm, A. Saulino, and F.S. Collins. 1991. Construction of T-vector, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19:** 1154.
31. Kovalic, D., J.H. Kwak, and B. Weisblum. 1991. General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nucleic Acids Res.* **19:** 4650.
32. Holton, T.A. and M.W. Graham. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nuceic Acids Res.* **19:** 1156.
33. Stoker, A.W. 1990. Cloning of PCR products after defined cohesive termini are created with T4 DNA polymerase. *Nucleic Acids Res.* **18:** 4290.
34. Iwahana, H., N. Mizusawa, S. Ii, K. Yoshimoto, and M. Itakura. 1994. An end-trimming method to amplify adjacent cDNA fragments by PCR. *BioTechniques* **16:** 94–98.
35. Crowe, J.S., H.J. Cooper, M.A. Smith, M.J. Sims, D. Parker, and D. Gewert. 1991. Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. *Nucleic Acids Res.* **19:** 184.
36. Thweatt, R., S. Goldstein, and R.J.S. Reis. 1990. A universal primer mixture for sequence determination at the 3′ ends of cDNAs. *Anal. Biochem.* **190:** 314.
37. Eckert, K.A. and T.A. Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18:** 3739–3745.
38. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
39. Sarker, G., S. Kapelner, and S.S. Sommer. 1990. Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res.* **18:** 7465.

Seikagaku Corporation Exhibit 2004 - Page 101 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Downloaded from genome.cshlp.org on January 13, 2023 - Published by Cold Spring Harbor Laboratory Press



On beyond classic RACE (rapid amplification of cDNA ends).

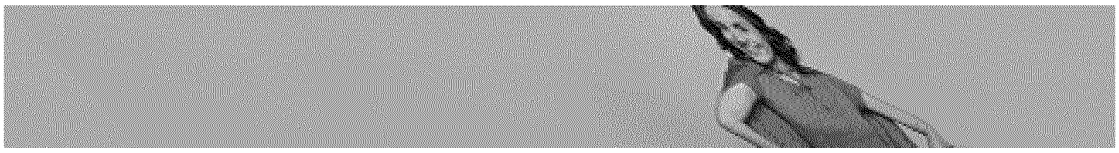
M A Frohman

Genome Res. 1994 4: S40-S58

References This article cites 36 articles, 5 of which can be accessed free at:
http://genome.cshlp.org/content/4/1/S40.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Copyright © Cold Spring Harbor Laboratory Press

cDNA race kit

Bilder | Videos | News | Bücher | Maps | Flüge | Finanzen | Alle Filter | Werkzeug

Beliebige Sprache | Vor 18. Okt. 2016 | Alle Ergebnisse | Zurücksetzen

Gesponsert

sigmaaldrich.com
https://www.sigmaaldrich.com

RACE Kit, 2nd Generation sufficient for 10 reactions |...

Order top products for your research available from Merck today. Tools for Research. Brands: Millipore, Sigma-Aldrich, Supelco, SAFC, Milli-Q.

Thermo Fisher Scientific
https://www.thermofisher.com › p...



5' RACE System for Rapid Amplification of cDNA Ends, ...

22.08.2015 — The 5 **RACE** System, Version 2.0 is suitable for rapid amplification of **cDNA** ends (**RACE**) (1) and anchored PCR (2) between a defined point in the mRNA and the ...

Thermo Fisher Scientific
https://www.thermofisher.com › p...



3' RACE System for Rapid Amplification of cDNA Ends

20.08.2015 — 3 **RACE** System is suitable for rapid amplification of **cDNA** ends (**RACE**) (1-3) and anchored PCR between a defined point within mRNA and the 3 poly(A) end ...

Sigma-Aldrich
https://www.sigmaaldrich.com › pcr

5'/3' RACE Kit, 2nd Generation Troubleshooting

12.09.2015 — Rapid amplification of **cDNA** ends (**RACE**) is used to rapidly obtain full-length **cDNA** from partially known sequences. The 5'/3' **RACE kit** (Product No.

Fisher Scientific
https://www.fishersci.com › shop

Invitrogen 5' RACE System for Rapid Amplification of cDNA ...

06.12.2004 — The 5' **RACE** System provides a set of prequalified reagents intended for synthesis of first-strand **cDNA**, purification of first-strand products, homopolymeric ...
1.128,65 $

Molecular Cloning Lab
https://www.mclab.com › Ab-clo...

Ab Cloning 5'-RACE Kit

10.12.2014 — The Ab cloning **RACE kit** provides a novel method for performing both 5' and 3' rapid amplifications of **cDNA** ends (**RACE**) with optimized primers for cloning ...

다카라
http://www.takara.co.kr › file › manual › pdf

SMART RACE cDNA Amplification Kit User Manual

06.06.2007 — SMART **RACE cDNA** amplification is a flexible tool—many researchers use this **kit** in place of conventional **kits** to amplify just the 5' or 3' end of a particular ...
41 Seiten

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › P...

Targeted rapid amplification of cDNA ends (T-RACE)

von NI Bower · 2010 · Zitiert von: 30 — Many of the commercially available **kits** offer protocols that allow the amplification of 5' **cDNA** ends of many genes from a single **cDNA** synthesis...

基礎生物学研究所
https://www.nibb.ac.jp › 4.5.htm

4.5 RACE

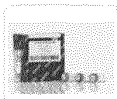
17.07.2005 — The GeneRacer **kit** provides a method to obtain full-length 5' and 3' ends of **cDNA** using known **cDNA** sequence. The **kit** ensures the amplification of only full- ...

issuu
https://issuu.com › docs › 633779...

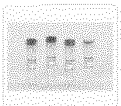


Kit SMARTer®RACE 5'/3' by AllScience LLC

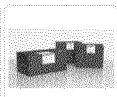
22.08.2016 — With SMART **cDNA** synthesis technology and a streamlined method, SMARTer **RACE kits** provide a rapid, easy way to analyze these variable and important nucleotide ...



Omniscript RT Kit (50)
321,00 €
www.qiagen.com
+36,00 € Ver...
Von Google



Reverse Transcription ...
137,00 €
www.qiagen.com
+36,00 € Ver...
Von Google



TaqIT DNA Polymerase...
581,00 €
www.qiagen.com
+36,00 € Ver...
Von Google

Invitrogen™
life technologies

Instruction Manual

5´ RACE System for Rapid Amplification of cDNA Ends, Version 2.0

Catalog no. 18374-058

Version E
6 December 2004
50327

Table of Contents

1. Notices to Customer 1
- 1.1 Important Information 1
- 1.2 Precautions 1
- 1.3 Limited Label Licenses 1

2. Overview 2
- 2.1 Principles of RACE 2
- 2.2 Summary of the 5' RACE System 2
  - 2.2.1 Isolation of RNA 4
  - 2.2.2 Design of 5' RACE Primers 4
  - 2.2.3 First Strand cDNA Synthesis from Total RNA 5
  - 2.2.4 Removal of RNA Template 5
  - 2.2.5 Purification of First Strand Product 6
  - 2.2.6 Homopolymeric Tailing of cDNA 6
  - 2.2.7 Amplification of Target cDNA 6
  - 2.2.8 Cloning of 5' RACE Amplification Products 9

3. Methods 10
- 3.1 Components and Storage 10
- 3.2 Additional Materials Required 11
- 3.3 Performance and Limitations of Procedures 11
- 3.4 Advance Preparations 11
  - 3.4.1 Isolation of Total RNA 11
  - 3.4.2 Design of the Gene-Specific Primers 12
  - 3.4.3 1X Wash Buffer for S.N.A.P. Procedure 13
  - 3.4.4 70% Ethanol Wash for S.N.A.P. Procedure 13
- 3.5 First Strand cDNA Synthesis 13
- 3.6 S.N.A.P Column Purification of cDNA 14
- 3.7 TdT Tailing of cDNA 14
- 3.8 PCR of dC-Tailed cDNA 15
- 3.9 Nested Amplification 16

4. Interpretation of Results 17

5. Troubleshooting Guide 18
- 5.1 Testing the 5' RACE System Using the Control RNA and DNA 18
- 5.2 Control First Strand cDNA Synthesis 21
- 5.3 S.N.A.P Column Purification of Control cDNA 21
- 5.4 TdT Tailing of the Control First Strand cDNA 22
- 5.5 PCR of cDNA, Tailed cDNA and Control DNA 22
- 5.6 General Troubleshooting Guidelines 24
- 5.7 Minimizing RNase Contamination 29
- 5.8 T_m Values for 5' RACE and Control Primers 29

Seikagaku Corporation Exhibit 2004 - Page 106 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Table of Contents

6. Additional Protocols 31
- 6.1 5' RACE of G:C Rich cDNA 31
  - 6.1.1 Alternative Protocol for First Strand cDNA Synthesis of Transcripts with High GC Content 31
- 6.2 dA-Tailing of cDNA 32
  - 6.2.1 Second-Strand Synthesis 33
  - 6.2.2 PCR of cDNA 33
- 6.3 DNase I Digestion of RNA Preparation 34
- 6.4 Lithium Chloride Purification of RNA Preparation 34
- 6.5 Amplification from an Agarose Plug 35
- 6.6 Proteinase K Treatment of 5' RACE Product 35
- 6.7 Cloning Using T4 DNA Polymerase 35

7. References 37

8. Related Products 40

Figures

1 Overview of the 5' RACE System Procedure 3
2 Sequences of the 5' RACE Primers 4
3 Effect of GSP1 on 5' RACE Efficiency 7
4 Sequences of the Control Primers 18
5 Agarose Gel Analysis of 5' RACE Products Using Control RNA 19
6 Flow Diagram for Troubleshooting the 5' RACE System Using the Control RNA and DNA 20
7 Sequence of the 3' RACE Adapter Primer 31
9 Cloning of 5' RACE Products using T4 DNA Polymerase 36

Tables

1 T_m Values of 5' RACE and Control Primers 30
2 Protruding Ends Compatible with *Acc* I or *Not* I 36

Sephadex® is a registered trademark of Amersham Pharmacia Biotech A.B.

Seikagaku Corporation Exhibit 2004 - Page 107 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Notices to Customer

1

1.1 Important Information

This product is authorized for laboratory research use only. The product has not been qualified or found safe and effective for any human or animal diagnostic or therapeutic application. Uses for other than the labeled intended use may be a violation of applicable law.

1.2 Precautions

Warning: This product contains hazardous reagents. It is the end-user's responsibility to consult the applicable MSDS(s) before using this product. Disposal of waste organics, acids, bases, and radioactive materials must comply with all appropriate federal, state, and local regulations.

1.3 Limited Label Licenses

Limited Use Label License No. 4: Products for PCR that include no rights to perform PCR

This product is optimized for use in the Polymerase Chain Reaction (PCR) covered by patents owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche, Ltd. ("Roche"). No license under these patents to use the PCR process is conveyed expressly or by implication to the purchaser by the purchase of this product. A license to use the PCR process for certain research and development activities accompanies the purchase of certain reagents from licensed suppliers such as Invitrogen, when used in conjunction with an Authorized Thermal Cycler, or is available from Applied Biosystems. Further information on purchasing licenses to practice the PCR process may be obtained by contacting the Director of Licensing at Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404 or at Roche Molecular Systems, Inc., 1145 Atlantic Avenue, Alameda, California 94501.

Limited Label License No. 5: SuperScript™ Reverse Transcriptase

The purchase of this product conveys to the buyer the non-transferable right to use the purchased amount of the product and components of the product in research conducted by the buyer (whether the buyer is an academic or for-profit entity). The buyer cannot sell or otherwise transfer (a) this product (b) its components or (c) materials made using this product or its components to a third party or otherwise use this product or its components or materials made using this product or its components for Commercial Purposes. The buyer may transfer information or materials made through the use of this product to a scientific collaborator, provided that such transfer is not for any Commercial Purpose, and that such collaborator agrees in writing (a) not to transfer such materials to any third party, and (b) to use such transferred materials and/or information solely for research and not for Commercial Purposes. Commercial Purposes means any activity by a party for consideration and may include, but is not limited to: (1) use of the product or its components in manufacturing; (2) use of the product or its components to provide a service, information, or data; (3) use of the product or its components for therapeutic, diagnostic or prophylactic purposes; or (4) resale of the product or its components, whether or not such product or its components are resold for use in research. Invitrogen Corporation will not assert a claim against the buyer of infringement of patents owned by Invitrogen and claiming this product based upon the manufacture, use or sale of a therapeutic, clinical diagnostic, vaccine or prophylactic product developed in research by the buyer in which this product or its components was employed, provided that neither this product nor any of its components was used in the manufacture of such product. If the purchaser is not willing to accept the limitations of this limited use statement, Invitrogen is willing to accept return of the product with a full refund. For information on purchasing a license to this product for purposes other than research, contact Licensing Department, Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, California 92008. Phone (760) 603-7200. Fax (760) 602-6500.

Limited Use Label License No. 43: Universal Amplification Primer

This product is the subject of U.S. Patent No. 5,334,515 owned by Invitrogen Corporation. The purchase of this product conveys to the buyer the non-transferable right to use the purchased amount of the product and components of the product in research conducted by the buyer (whether the buyer is an academic or for-profit entity). The buyer cannot sell or otherwise transfer (a) this product (b) its components or (c) materials made using this product or its components to a third party or otherwise use this product or its components or materials made using this product or its components for Commercial Purposes. The buyer may transfer information or materials made through the use of this product to a scientific collaborator, provided that such transfer is not for any Commercial Purpose, and that such collaborator agrees in writing (a) to not transfer such materials to any third party, and (b) to use such transferred materials and/or information solely for research and not for Commercial Purposes. Commercial Purposes means any activity by a party for consideration and may include, but is not limited to: (1) use of the product or its components in manufacturing; (2) use of the product or its components to provide a service, information, or data; (3) use of the product or its components for therapeutic, diagnostic or prophylactic purposes; or (4) resale of the product or its components, whether or not such product or its components are resold for use in research. Invitrogen Corporation will not assert a claim against the buyer of infringement of the above patent based upon the manufacture, use or sale of a therapeutic, clinical diagnostic, vaccine or prophylactic product developed in research by the buyer in which this product or its components was employed, provided that neither this product nor any of its components was used in the manufacture of such product. If the purchaser is not willing to accept the limitations of this limited use statement, Invitrogen is willing to accept return of the product with a full refund. For information on purchasing a license to this product for purposes other than research, contact Licensing Department, Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, California 92008. Phone (760) 603-7200. Fax (760) 602-6500.

Limited Use Label License No. 44: 5' RACE Products

This product is the subject of one or more of U.S. Patent Nos. 5,578,467; 5,334,515; and 5,869,251 owned by Invitrogen Corporation. The purchase of this product conveys to the buyer the non-transferable right to use the purchased amount of the product and components of the product in research conducted by the buyer (whether the buyer is an academic or for-profit entity). The buyer cannot sell or otherwise transfer (a) this product (b) its components or (c) materials made using this product or its components to a third party or otherwise use this product or its components or materials made using this product or its components for Commercial Purposes. The buyer may transfer information or materials made through the use of this product to a scientific collaborator, provided that such transfer is not for any Commercial Purpose, and that such collaborator agrees in writing (a) to not transfer such materials to any third party, and (b) to use such transferred materials and/or information solely for research and not for Commercial Purposes. Commercial Purposes means any activity by a party for consideration and may include, but is not limited to: (1) use of the product or its components in manufacturing; (2) use of the product or its components to provide a service, information, or data; (3) use of the product or its components for therapeutic, diagnostic or prophylactic purposes; or (4) resale of the product or its components, whether or not such product or its components are resold for use in research. Invitrogen Corporation will not assert a claim against the buyer of infringement of the above patents based upon the manufacture, use or sale of a therapeutic, clinical diagnostic, vaccine or prophylactic product developed in research by the buyer in which this product or its components was employed, provided that neither this product nor any of its components was used in the manufacture of such product. If the purchaser is not willing to accept the limitations of this limited use statement, Invitrogen is willing to accept return of the product with a full refund. For information on purchasing a license to this product for purposes other than research, contact Licensing Department, Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, California 92008. Phone (760) 603-7200. Fax (760) 602-6500.

Seikagaku Corporation Exhibit 2004 - Page 108 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

2 Overview

2.1 Principles of RACE

Rapid Amplification of cDNA Ends (RACE) is a procedure for amplification of nucleic acid sequences from a messenger RNA template between a defined internal site and unknown sequences at either the 3' or the 5' -end of the mRNA (1). This methodology of amplification with single-sided specificity has been described by others as "one-sided" PCR (2) or "anchored" PCR (3). In general, PCR amplification of relatively few target molecules in a complex mixture requires two sequence-specific primers that flank the region of sequence to be amplified (4,5). However, to amplify and characterize regions of unknown sequences, this requirement imposes a severe limitation (3). 3' and 5' RACE methodologies offer possible solutions to this problem.

3' RACE takes advantage of the natural poly(A) tail in mRNA as a generic priming site for PCR amplification. In this procedure, mRNAs are converted into cDNA using reverse transcriptase (RT) and an oligo-dT adapter primer. Specific cDNA is then directly amplified by PCR using a gene-specific primer (GSP) that anneals to a region of known exon sequences and an adapter primer that targets the poly(A) tail region. This permits the capture of unknown 3'-mRNA sequences that lie between the exon and the poly(A) tail.

5' RACE, or "anchored" PCR, is a technique that facilitates the isolation and characterization of 5' ends from low-copy messages. The method has been reviewed by both Frohman (6,8) and Loh (7). Although the precise protocol varies among different users, the general strategy remains consistent. First strand cDNA synthesis is primed using a gene-specific antisense oligonucleotide (GSP1). This permits cDNA conversion of specific mRNA, or related families of mRNAs, and maximizes the potential for complete extension to the 5' -end of the message. Following cDNA synthesis, the first strand product is purified from unincorporated dNTPs and GSP1. TdT (Terminal deoxynucleotidyl transferase) is used to add homopolymeric tails to the 3' ends of the cDNA. In the original protocol, tailed cDNA is then amplified by PCR using a mixture of three primers: a nested gene-specific primer (GSP2), which anneals 3' to GSP1; and a combination of a complementary homopolymer-containing anchor primer and corresponding adapter primer which permit amplification from the homopolymeric tail. This allows amplification of unknown sequences between the GSP2 and the 5'-end of the mRNA.

RACE procedures have been used for amplification and cloning of rare mRNAs that may escape, or prove challenging for, conventional cDNA cloning methodologies (7). Additionally, RACE may be applied to existing cDNA libraries (9). Random hexamer-primed cDNA has also been adapted to 5' RACE for amplification and cloning of multiple genes from a single first strand synthesis reaction (10). Products of RACE reactions can be directly sequenced without any intermittent cloning steps (11,12), or the products can be used for the preparation of probes (13). Products generated by the 3' and 5' RACE procedures may be combined to generate full-length cDNAs (6,13). Lastly, the RACE procedures may be utilized in conjunction with exon trapping methods (14) to enable amplification and subsequent characterization of unknown coding sequences.

2.2 Summary of the 5' RACE System

The 5' RACE System is a set of prequalified reagents intended for synthesis of first strand cDNA, purification of first strand products, homopolymeric tailing, and preparation of target cDNA for subsequent amplification by PCR. Control RNA, DNA, and primers are provided for monitoring system performance. The procedure is summarized in figure 1.

Seikagaku Corporation Exhibit 2004 - Page 109 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

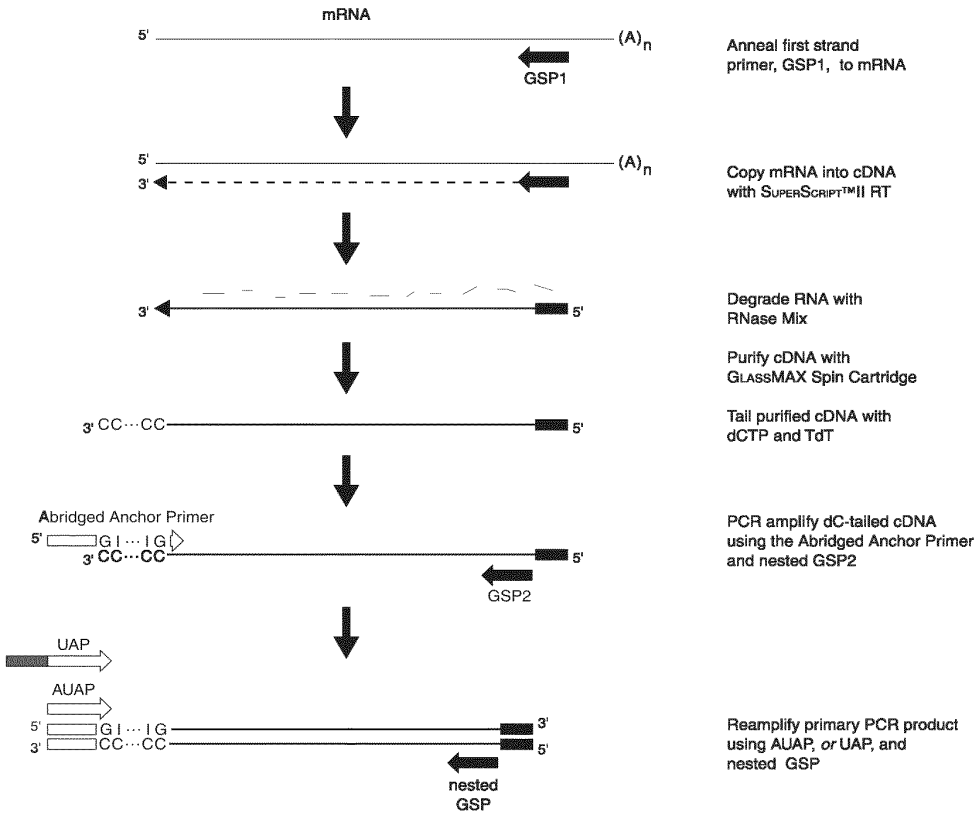


Figure 1. Overview of the 5' RACE Procedure.

First strand cDNA is synthesized from total or poly(A)$^+$ RNA using a gene-specific primer (GSP1) that the user provides and SuperScript™ II, a derivative of Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) with reduced RNase H activity. After first strand cDNA synthesis, the original mRNA template is removed by treatment with the RNase Mix (mixture of RNase H, which is specific for RNA:DNA heteroduplex molecules, and RNase T1). Unincorporated dNTPs, GSP1, and proteins are separated from cDNA using a S.N.A.P. Column. A homopolymeric tail is then added to the 3'-end of the cDNA using TdT and dCTP. Since the tailing reaction is performed in a PCR-compatible buffer, the entire contents of the reaction may be directly amplified by PCR without intermediate organic extractions, ethanol precipitations, or dilutions. PCR amplification is accomplished using *Taq* DNA polymerase, a nested, gene-specific primer (GSP2, designed by the user) that anneals to a site located within the cDNA molecule, and a novel deoxyinosine-containing anchor primer (patent pending) provided with the system.

Following amplification, 5' RACE products can be cloned into an appropriate vector for subsequent characterization procedures, which may include sequencing, restriction mapping, preparation of probes to detect the genomic elements associated with the cDNA of interest, or *in vitro* RNA synthesis. The Abridged Anchor Primer (AAP), Abridged Universal Amplification Primer (AUAP), Anchor Primer (AP) [available separately], and Universal Amplification Primer (UAP) include recognition sequences for *Mlu* I, *Sal* I, and *Spe* I to facilitate restriction endonuclease cloning of RACE products (figure 2).

Seikagaku Corporation Exhibit 2004 - Page 110 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

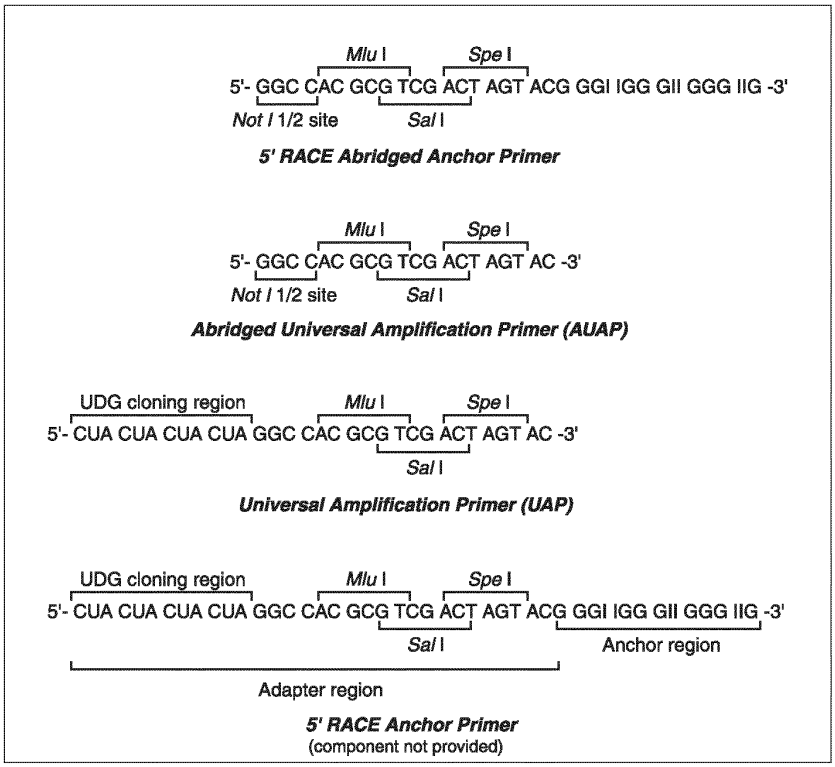


Figure 2. Sequence of the 5' RACE Primers.

5' RACE products may also be used for procedures that do not require an intermittent cloning step such as dsDNA cycle sequencing (12,18) or probe preparation (13). However, additional rounds of PCR using the AUAP, or UAP, in conjunction with either progressively nested GSPs or size-selected products from the initial PCR, may be required to confer an adequate level of specificity to the process to permit direct characterization of RACE products.

Details of the individual steps of 5' RACE are discussed below. Review this information carefully before beginning. The RNA isolation, design of primers, and the amplification protocols are most important for optimal results.

2.2.1 Isolation of RNA

The quality of the RNA dictates the maximum amount of sequence information that can be converted into cDNA. Thus, it is important to optimize the isolation of RNA (19), and to prevent introduction of RNases and inhibitors of RT (20, 52). The guanidine isothiocyanate/acid-phenol method, originally described by Chomzynski and Sacchi (21), is the recommended method for RNA isolation. See section 3.4.1.

2.2.2 Design of 5' RACE Primers

The sensitivity and the specificity of the first strand synthesis and the subsequent PCR depend upon good primer design. A minimum of two antisense gene-specific primers (GSP) are **required** for 5' RACE and must be supplied by the user.

1. GSP1 primes first-strand cDNA synthesis (figure 1). Design this primer to anneal at least 300 bp from the mRNA 5'-end so that the cDNA can be easily purified using a S.N.A.P. column.
2. A second nested primer, GSP2, that anneals to sequences located 3' (with respect to cDNA not mRNA) of GSP1 is required for PCR. GSP2 can anneal immediately adjacent to GSP1 or at sequences located further upstream of GSP1 within the cDNA product. Other sequentially nested GSPs may be required depending on the efficiency and specificity of the primary PCR.

Seikagaku Corporation Exhibit 2004 - Page 111 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

In general, these primers should be highly specific for their target sequences, able to form stable duplexes with their target sequences, and free of secondary structure. The key rules for primer design are discussed in section 3.4.2 (as well as in references 22-26). The primers provided in this system were carefully designed for successful 5' RACE.

The anchor primers contain 3' sequence complementary to the homopolymeric tail and additional 5' sequence that encodes an adapter region, comprised of restriction endonuclease sites and other functional sequences which facilitate cloning and characterization of 5' RACE products. Normally, homopolymer primers create melting temperatures that are either higher [poly (dG)•poly (dC)] or lower [poly (dA)•poly (dT)] than a typical GSP. They also can have poor specificity that can lead to mispriming at internal sequences. To minimize these problems, our anchor primers (figure 2, patent pending) were designed with the selective placement of deoxyinosine residues in the poly (dG) portion. This design eliminates the need to use the mixtures of anchor and adapter primers described in the original method (6,7).

Note: While the anchor primers enable efficient amplification of many target sequences, they may not be an idealized solution for all 5' RACE applications. It is possible that the anchor primer may anneal at certain gene-specific sequences. Therefore, as in any RACE procedure, specificity of the anchor primer for the oligo-dC tail should be tested by performing amplification reactions with cDNA subjected to dC-tailing both in the presence of and absence of TdT.

Deoxyinosine has the capacity to base-pair with all four bases; however, it does so with varying affinities. The order of stabilities for the different combinations, from greatest to least stable, reported by Martin *et al.* are as follows: I:C, I:A, I:T, and I:G. I:C pairs were found to be slightly less stable than A:T pairs (27). The selective placement of deoxyinosine residues in the 3' region of the anchor primer maintains low stability on the primer's 3'-end (ΔG = -8.2 kCal/mol) and creates a melting temperature (T_m) for the 16-base anchor region (66.6°C) which is comparable to that of a typical 20-mer primer with 50% GC content (22,23). This maximizes specific priming from the oligo-dC tail, minimizes priming at internal C-rich regions of the cDNA, and establishes a relationship of a "balanced" T_m for the anchor region to that of GSP2, which is required for efficient PCR (6,7).

The Abridged Universal Amplification Primer (AUAP) and Universal Amplification Primer (UAP) are used to reamplify primary 5' RACE PCR products in applications such as nested PCR or enrichment of RACE products for cloning. The AUAP contains a restriction endonuclease site sequence (adapter region) homologous to the adapter region of the anchor primer.

The UAP is composed of the same adapter region as the AUAP plus a dUMP-containing sequence at the 5'-end of the primer required for uracil DNA glycosylase (UDG)-mediated cloning of 5' RACE products. The original 5' RACE Anchor Primer is available separately for applications that require UDG cloning of 5' RACE products directly from the primary PCR. The UAP, the 5' RACE Anchor Primer or any dUMP-containing primer should not be used to prime DNA synthesis with any archaeobacterial polymerase (*Pfu* DNA Polymerase, *Pwo* DNA Polymerase, etc.), including long PCR enzyme mixtures (28,29), because dUMP inhibits these polymerase activities.

2.2.3 First Strand cDNA Synthesis from Total RNA

The capture of mRNA 5'-ends is dependent on complete cDNA synthesis. The use of RNase H$^-$ RT for first strand synthesis results in greater full-length cDNA synthesis and higher yields of first strand cDNA than obtained with other RTs (20,30). SuperScript™ II RT has been engineered to retain the full DNA polymerase activity found in M-MLV RT (31). The enzyme exhibits increased thermal stability and may be used at temperatures up to 50°C. Because SuperScript™ II RT is not inhibited significantly by ribosomal and transfer RNA, it may be used effectively to synthesize first strand cDNA from a total RNA preparation. The RNA template is removed from the first strand cDNA product as described below.

2.2.4 Removal of RNA Template by RNase Mix

After cDNA synthesis, RNase Mix, a mixture of RNase H and RNase T1, is used to degrade the RNA. The digestion is performed following thermal inactivation of the RT in order to reduce the potential for hairpin-primed second-strand synthesis (catalyzed by RT) which can obscure the accessibility of the cDNA ends to TdT. Template RNA in the cDNA:RNA hybrid is degraded by RNase H and the single-stranded RNAs are degraded by RNase T1. This eliminates possible renaturation of template RNA to cDNA. TdT does not use RNA as a substrate (32); however, RNA may inhibit tailing and subsequent PCR of the cDNA (33). Use of the RNase Mix upon completion of first strand synthesis is cru-

Seikagaku Corporation Exhibit 2004 - Page 112 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

cial to the efficiency of the tailing reaction because TdT exhibits a marked preference for single-stranded substrates (34,35).

2.2.5 Purification of First Strand Product

Excess nucleotides and GSP1 must be removed from the first strand product. Otherwise, residual GSP1 will be tailed by TdT and will compete for Abridged Anchor Primer during PCR (figure 3). Because of the large amounts of GSP1 relative to cDNA product, a stringent purification procedure is required (6,7,36). The S.N.A.P. column procedure, adapted from a method described by Vogelstein and Gillespie (37), provides a rapid and efficient means to purify first strand product. In the presence of the chaotropic agent, sodium iodide, cDNA >200 bases are bound to the silica-based membrane. Buffer components, dNTPs, enzymes, and oligonucleotides remain in solution and are removed by centrifugation with the effluent. Residual impurities and sodium iodide are removed by passing several volumes of 1X wash buffer followed by a 70% ethanol rinse through the S.N.A.P. column. Purified cDNA is recovered in distilled water and may be used directly in the TdT tailing reaction.

2.2.6 Homopolymeric Tailing of cDNA

TdT tailing creates the abridged anchor primer binding site on the 3'-end of the cDNA. Efficient tailing is necessary to provide:

1. A high proportion of tailed cDNA molecules for efficient amplification of first strand products.
2. Homopolymeric tails of sufficient length to allow the primer to anneal.
3. Homopolymeric tails of uniform length to produce a homogeneous amplification product.
4. A buffer compatible with the PCR buffer system.

The 5' RACE System tailing reaction has been optimized to meet these criteria. The 5' RACE System uses a tailing buffer [10 mM Tris-HCl (pH 8.4), 25 mM KCl and 1.5 mM $MgCl_2$] supplemented with 200 µM dCTP for homopolymeric tailing of first strand cDNA. The tailing reaction is highly sensitive to the concentration of each buffer component.

Concentrations of $MgCl_2$ in excess of 1.5 mM may significantly inhibit both the length of the tail and the percentage of molecules tailed. In general, components such as Tris buffers and salts have been reported to be inhibitory to TdT (32), and $CoCl_2$ has classically been chosen over $MgCl_2$ as the optimal divalent cation for tailing reactions (35). However, careful manipulation of buffers containing these components has been shown to produce results that are highly effective for 5' RACE (36).

Double-stranded 3' termini and hairpin structures may significantly impair homopolymeric tailing of cDNA; therefore, a brief denaturation procedure prior to tailing is used to disrupt any secondary structure in the cDNA.

The choice of nucleotide for homopolymeric tailing has been a subject of debate. Each nucleotide offers unique advantages and disadvantages. The 5' RACE System uses dC-tailing to complement our unique Abridged Anchor Primer (see section 2.2.2). dA-tailing permits the use of the same oligo-dT anchor primer for both 5' and 3' RACE procedures. However, because A:T base pairs are less stable than G:C base pairs, longer stretches of dAs or dTs are required for priming as compared to dGs or dCs.

2.2.7 Amplification of Target cDNA

Successful 5' RACE is extremely dependent on the efficiency and specificity of the PCR. Optimal conditions for amplification are dependent on the nature of each particular primer and target sequence used. Alteration of the magnesium ion, dNTP, or primer concentration, as well as the thermal cycling protocol, may be required. The optimal **free** magnesium concentration for efficient amplification is reported to be between 0.7 and 0.8 mM (38). Since magnesium can bind deoxynucleoside triphosphates, this factor is affected by both primer and dNTP concentration. In general, lower concentrations of dNTP (50 to 200 µM), $MgCl_2$ (1 to 1.5 mM), and primer (0.1 to 0.2 µM) promote higher fidelity and

Seikagaku Corporation Exhibit 2004 - Page 113 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

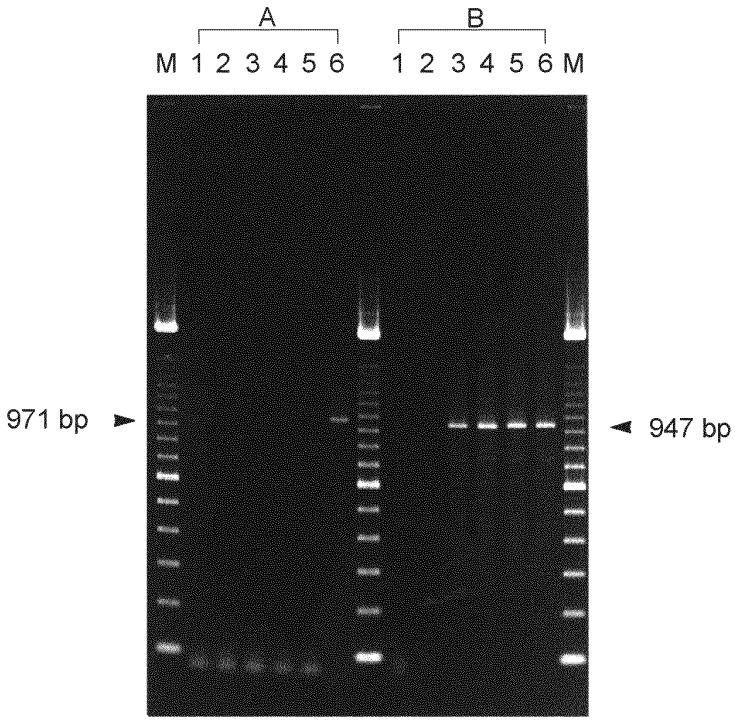


Figure 3. Effect of GSP1 on 5' RACE. A large-scale preparation of control cDNA was synthesized using 100 µg of an *in vitro* transcribed RNA essentially as described in *Methods* with the amount of components scaled appropriately. RNA template was degraded using RNase Mix. The cDNA product was purified by phenol:chloroform extraction followed by Sephadex® G-150 chromatography. Purified product was quantified by A_{260} and diluted in TE buffer. 4 x 10^5 copies of cDNA were tailed as described in the presence of 1, 0.1, 0.01, 0.001, or 0 pmoles of GSP1, lanes 1-6 respectively. One-fourth (1 x 10^5 copies of cDNA) of each reaction was amplified by PCR using Anchor Primer in combination with either (A) the control GSP1 or (B) the control GSP2 (see figure 4). One tenth (5 µl) of each PCR was analyzed by 1.5% agarose gel in 1X TBE stained with 0.5 µg/ml ethidium bromide. Lane M, Gibco BRL 100 bp DNA Ladder.

Seikagaku Corporation Exhibit 2004 - Page 114 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

specificity (39). Higher nucleotide concentrations, however, can be used to improve product yield as well as to promote 3'-terminal T-mismatches (40). This factor may warrant consideration if a degenerate oligonucleotide is used as GSP2. Additionally, since a degenerate primer represents a composite of many different priming sequences, higher primer concentrations ($\geq$1 µM) are generally required.

Typical thermal cycling parameters are provided in the protocols. However, optimal conditions depend not only on the primers and template, but also the type of PCR tube as well as the thermal cycler.

There are four times and temperatures that must be considered:

1) Preamplification denaturation of the template DNA (PAD)

2) Denaturation of product DNA at the beginning of each cycle

3) Annealing of the primers to the denatured DNA

4) Extension of the primers by the polymerase

Steps 2-4 constitute a cycle and are repeated usually 30-35 times followed by a final extension time of 5-10 min and then a holding temperature of 5°C.

Many PCR protocols use a PAD step of 3 to 5 min. However, an extended PAD is not usually necessary and may impair the ability to amplify longer sequences (41). The denaturation temperature and time should be sufficient to completely separate target strands, yet minimized to reduce deamination and depurination of target DNA. For thin-walled tubes in thermal cyclers which use the sample temperature (or calculated sample temperature) to control temperature cycling, a denaturation time of 10 s to 15 s at 94°C is adequate. Likewise, an annealing time of 20 s to 30 s is usually ample. In contrast, PCR in conventional 0.5-ml microcentrifuge tubes may require 1 min for complete denaturation and 30 s to 1 min for annealing. Optimal annealing temperature is dependent upon the thermodynamic properties of the primers (see section 3.4.2*j*). However, well-designed primers, *i.e.* primers with unstable 3'-ends ($\Delta G > -9$ kCal/mol), can function effectively in PCR over a broad range of annealing temperatures (41). A general rule for extension time is to allow 1 min for every 1 kb of target sequence. If primer T_ms are ≥68°C, a two step PCR, which cycles between denaturation at 94°C and combined annealing and extension at 68°C can be used. For a detailed discussion of parameters affecting PCR, please refer to Innis and Gelfand (42) or Saiki (38, 43).

Nonspecific annealing and extension of primers prior to the initial denaturation step of the PCR process may adversely affect the efficiency and specificity of amplification. These artifacts can be minimized by using the "hot start" technique (44,45) which requires the addition of either *Taq* DNA polymerase, dNTPs, or $MgCl_2$ after reactions have been equilibrated at 75°C to 80°C. For many applications it is sufficient to assemble the reactions on ice, in thin-walled PCR tubes and directly transfer the tubes to a thermal cycler equilibrated to the initial denaturation temperature, 94°C. This is the procedure used in section 3.9.

Amplification of a target cDNA synthesized with the 5' RACE System requires priming with two oligonucleotides. The Abridged Anchor Primer, which is specific for the oligo-dC tail added by TdT, serves as the sense primer. The antisense primer (GSP2), provided by the user, should anneal to an internal (nested) site within the cDNA sequence (with respect to the primer used for first strand synthesis, GSP1) and may include sequence elements that facilitate subsequent cloning steps (section 4.2).

Use of a nested GSP2 is essential for effective PCR (6,7,36). This not only adds a level of specificity to the process, but it prevents "primer-dimer" amplification of residual GSP1 that may carry through the cDNA purification procedure. This effect is illustrated in figure 3. Residual GSP1, which is subsequently tailed by TdT, is copied by extension of the anchor primer during PCR. This results in amplification of the tailed GSP1 sequence and blocks amplification of cDNA.

Seikagaku Corporation Exhibit 2004 - Page 115 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

2.2.8 Cloning 5' RACE Amplification Products

Conventional cloning methods that typically involve end-repair and blunt-end cloning can be problematic for amplified products (18,46,47). An alternative is a rapid and efficient cloning method involving the use of UDG (48-50). This method requires that the user design a nested GSP2 containing a 5'-$(CAU)_4$ sequence. Incorporation of dUMP into the nested GSP2 may be accomplished with minimal expense on most automated synthesizers or by ordering through Invitrogen's Custom Primers (see section 3.4.2).

An alternative to conventional cloning methods uses the 3' to 5' exonuclease activity of T4 DNA polymerase as the basis for cloning as described by Stoker (51). In this procedure, PCR products from the primary PCR with the Abridged Anchor Primer, or nested amplification reaction primed with the AUAP, are treated with T4 DNA polymerase to generate a *Not* I 5' overhang.

Another approach to cloning is to digest the 5' RACE product using one of the restriction endonuclease sites designed into the AUAP (see figure 2) (1). The user may also design unique restriction sites into the GSP, exploit a site present in the cDNA sequence or end-repair the 5' RACE product prior to restriction-endonuclease digestion (46).

Seikagaku Corporation Exhibit 2004 - Page 116 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

3 Methods

3.1 Components and Storage

Note: The 5' RACE system does not include *Taq* DNA polymerase or the reagents required for cloning.

The components of the 5' RACE System are as follows. Sufficient material is provided for 10 reactions. One reaction prepares specific cDNA from 1-5 µg of a total RNA or 50-500 ng poly $(A)^+$ RNA preparation for amplification by anchored PCR. The amount of RNA will vary depending on the application. Control RNA, DNA, and primers are included to verify the performance of the system and may be added to experimental RNA preparations to monitor the efficiency of each step or to troubleshoot potential problems.

Note: The 10X PCR buffer does not contain $MgCl_2$. Therefore, $MgCl_2$ must be added to the first strand reaction mix.

Component	Volume	Storage
Reagents:		
10X PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl]	500 µl	–20°C
25 mM $MgCl_2$	500 µl	–20°C
10 mM dNTP mix [10 mM each dATP, dCTP, dGTP, dTTP]	100 µl	–20°C
0.1 M DTT	100 µl	–20°C
SuperScript™ II Reverse Transcriptase (200 units/µl)	10 µl	–20°C
RNase mix	10 µl	–20°C
5X tailing buffer [50 mM Tris-HCl (pH 8.4), 125 mM KCl, 7.5 mM $MgCl_2$]	500 µl	–20°C
2 mM dCTP	50 µl	–20°C
terminal deoxynucleotidyl transferase	15 µl	–20°C
5' RACE abridged anchor primer (AAP, 10 µM)	80 µl	–20°C
universal amplification primer (UAP, 10 µM)	40 µl	–20°C
abridged universal amplification primer (AUAP, 10 µM)	40 µl	–20°C
DEPC-treated water	1.25 ml	–20°C
control gene-specific primer 1 (GSP1, 1 µM)	25 µl	–20°C
control nested gene-specific primer 2 (GSP2, 10 µM)	80 µl	–20°C
control PCR primer, gene-specific primer 3 (GSP3, 10 µM)	20 µl	–20°C
control DNA (2 x 10^4 copies/µl; ~0.1 pg/µl)	100 µl	–20°C
control RNA (50 ng/µl)	10 µl	–70°C
DNA Purification System:		
S.N.A.P. Columns	10 columns	4°C
Collection tubes	10 tubes	4°C
binding solution (6 M sodium iodide)	30 ml	4°C
Wash buffer concentrate	1 ml	4°C

Note: Do not freeze the DNA Purification System.

Seikagaku Corporation Exhibit 2004 - Page 117 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

3.2 Additional Materials Required

The following items are required for use with the 5' RACE System, but are not included.

- sterilized, RNase-free thin-walled 0.2 or 0.5-ml PCR tubes;
- automatic pipettes capable of dispensing 1 to 20 µl and 20 to 200 µl;
- sterilized, RNase-free disposable tips for automatic pipettes;
- disposable latex gloves;
- sterilized, distilled water;
- absolute ethanol;
- GSP1 (cDNA primer, user-defined);
- GSP2 (nested primer for PCR amplification of dC-tailed cDNA, user-defined, appropriately engineered);
- microcentrifuge capable of generating a relative centrifugal force of 13,000 x *g*;
- 37°C, 42°C, 65°C, and 70°C water baths or heat blocks (or use thermal cycler);
- *Taq* DNA Polymerase;
- programmable thermal cycler; and
- mineral oil (if necessary for your thermal cycler)

Note: All temperature incubations may be performed in an appropriately programmed thermal cycler. This eliminates the need for multiple fixed temperature baths or heat blocks.

3.3 Performance and Limitations of Procedures

The 5' RACE System has been functionally tested using the control RNA according to the protocols described in this manual using GIBCO BRL *Taq* DNA Polymerase. Following PCR, a distinct 711-bp band was visible by agarose gel electrophoresis and ethidium bromide staining. While the 5' RACE system provides a direct and reliable solution for the preparation of tailed cDNA, PCR with single-sided specificity remains highly challenging. Success with the system is extremely dependent on the efficiency and specificity of the PCR used to amplify your tailed cDNA. *Taq* DNA polymerase from other suppliers may not function as well in the buffers provided in this system. The PCR protocol, section 3.8, is intended as a starting point. Optimal amplification parameters for your target gene may vary.

Obtaining longer 5' RACE products, *i.e.* greater than 1 kb, adds an additional challenge to the procedure. The 5' RACE System has been used successfully with eLONGase®, Enzyme Mix, an enzyme system designed for amplification of long templates, to obtain an increased yield of amplification product as well as substantially increased length of 5' RACE products. The principle barrier to long(er) 5' RACE lies in the specificity and efficiency of PCR. Critical success factors include primer design, PCR optimization, and a systematic experimental strategy that includes amplification of primary PCR using nested, gene-specific primers. Truncated products can yield informative sequence data that can be applied in additional 5' RACE experiments as one walks toward the 5'-end.

3.4 Advance Preparations

Please review the advance preparation guidelines discussed in this section prior to starting to work with the 5' RACE System. To achieve optimal results, it is also recommended that you review Chapter 2 before using this system.

3.4.1 Isolation of Total RNA

One of the most important factors affecting the synthesis of substantially full-length cDNA is the isolation of intact RNA. Therefore, it is important to optimize the isolation of RNA and to prevent introduction of RNases (19) and inhibitors of RT such as guanidinium salts, SDS and EDTA (20,52). The recommended method of RNA isolation is the guanidine isothiocyanate/acid-phenol method originally described by Chomzynski and Sacchi (21).The TRIzol® Reagent method (53) is an improvement of the original single-step method of Chomczynski and Sacchi and can be used for the preparation of RNA from as little as 10^3 cells (54). Total RNA isolated with TRIzol® Reagent is undegraded and essentially free of protein and DNA contamination. To maintain intact RNA, an RNase-free environment (see section 5.7) is critical.

Seikagaku Corporation Exhibit 2004 - Page 118 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Total RNA may contain small amounts of genomic DNA that may be amplified along with the target cDNA. The presence of this double-stranded DNA is not likely to cause problems in 5' RACE, since it is inefficiently tailed prior to amplification. As a precaution, however, perform a control experiment without RT. Products generated in the absence of RT are of genomic origin. If your application requires removal of all genomic DNA from your RNA preparation, refer to section 6.3. Oligo(dT)-selection for poly(A)$^+$ RNA is typically not necessary although incorporating this step may facilitate the detection of rare mRNA transcripts.

3.4.2 Design of the Gene-Specific Primers

Efficient and specific PCR is highly dependent on effective primer design. This is especially true for RACE applications since the PCR is carried out with only a single GSP. No method of primer design can guarantee successful amplification, so all primers must be tested in PCR before they can be pronounced "good".

Primers for PCR (GSP2 AND GSP3): Effective primers form stable, highly specific duplexes with their target sequences, and are free of secondary structure such as hairpin loops and dimers (22,23,25,55). High stability, *i.e.,* G:C clamps, in the 5'- and central regions of the primer confer hybridization stability with the target sequence. Primers with unstable 3'-ends (ΔG > -9 kCal/mol) often result in higher specificity, since the potential to misprime at nontarget sites is reduced. Generally, this condition can be met by including no more than two G or C residues in the last five 3'-bases (22,23). Additionally, 3'-terminal complementarity should be minimized since primer-dimer artifacts may significantly reduce PCR efficiency. Therefore the nested amplification primer should be examined for dimer formation with the anchor primer, as well as itself. Computer algorithms that have been developed (22,56-58) often facilitate this analysis as well as secondary structure analysis.

The next important parameter for primers is the T_m (the temperature at which 50% of the primer and its complementary sequence are present in a duplex DNA molecule.) The T_m is necessary to establish an annealing temperature for PCR. The annealing temperature should be low enough to guarantee efficient annealing of the primer to the target, but high enough to minimize nonspecific binding. Since a single GSP is used in RACE, use as stringent an annealing temperature as possible to minimize amplification of nonspecific products. As a good starting point, choose an annealing temperature a few degrees below the estimated T_ms of the primer pair (see section 5.8 for 5' RACE primer T_ms). In practice, primers with T_ms between 60°C and 75°C usually can function effectively in 5' RACE. Ideally, the T_m of primers used for PCR should be closely matched (6).

Several methods are available to estimate the T_m of primers. These provide only estimates and the optimal annealing temperature must be established experimentally. When designing UDG cloning primers, consider dU residues as if they were dT residues for the calculations (see section 5.8.).

Primer for First Strand cDNA Synthesis (GSP1): For GSP1, follow the same rules used for PCR primer selection. The T_m should be appropriate for the relatively low temperature (42°C) of the cDNA synthesis reaction. A short primer (16 to 20 bases) facilitates efficient separation of GSP1 from cDNA product. Efficient recovery of cDNA from the S.N.A.P. column requires a product of at least 200 bases in length. Therefore, we recommend that GSP1 anneal to sequences located at least 300 bases from the mRNA 5'-end. Again, even well-designed primers may not function efficiently in priming first strand synthesis. For example, secondary structure of the mRNA at the annealing site may prevent efficient annealling or extension of the primer.

Primers for Subsequent Cloning: The user defined GSPs need to be compatible with the cloning method. Add the following to the 5'-end of the nested GSP:

for T4 DNA polymerase cloning 5'–CGA–3' (use with AUAP)

for restriction endonuclease cloning, an appropriate adapter region (59) is required.

It should be noted that in cases where only limited peptide sequence information is available, degenerate GSPs may be prepared. An alternative to the synthesis of degenerate primers is the substitution of dI residues at wobble-base positions (60-62).

Seikagaku Corporation Exhibit 2004 - Page 119 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

3.4.3 1X Wash Buffer for S.N.A.P. Procedure

Prior to using the system for the first time, a 1X wash buffer must be prepared from the wash buffer concentrate.

1. Pipette 1 ml of the wash buffer concentrate into a 50-ml graduated cylinder.
2. Add 18 ml of distilled water and 21 ml of absolute ethanol. Mix thoroughly.
3. Transfer to an appropriate-sized glass bottle. Cap and store at 4°C.

3.4.4 70% Ethanol Wash for S.N.A.P. Procedure

1. Add 35 ml of absolute ethanol and 15 ml of distilled water to a 50-ml graduated cylinder.
2. Transfer to an appropriate-sized glass bottle. Cap and store at 4°C.

3.5 First Strand cDNA Synthesis

A detailed protocol for using the control RNA in 5' RACE is provided in Chapter 5. You may wish to use this protocol to familiarize yourself with the procedure before attempting 5' RACE with your sample.

This procedure is designed to convert specific RNA sequence(s) from a background of 1-5 µg of total RNA into first strand cDNA. In general, 100 to 500 ng of total RNA should provide sufficient material for the amplification of low copy messages by 5' RACE (6). Although poly(A)$^+$ RNA may be used in this protocol to enrich for very rare messages, this level of purity is typically not necessary. If you wish to use the control RNA, refer to Sections 5.2 and 5.5 for proper procedure.

Note: Extreme care should be taken to avoid contamination of samples by RNase. (See section 5.7)

1. Add the following to a 0.5-ml microcentrifuge tube (or thin-walled PCR tube if using a thermal cycler):

Component	Amount
GSP1	2.5 pmoles (~10 to 25 ng)
sample RNA	1-5 µg
DEPC-treated water (or sterile, distilled water)	sufficient for a final volume of 15.5 µl

Note: If you choose to use placental RNase inhibitor in the reaction, it should be added *after* the addition of DTT. Appropriate volume adjustments should be made in step 1. See precautionary notes regarding the use of RNase inhibitor proteins in Chapter 5.

2. Incubate the mixture 10 min at 70°C to denature RNA. Chill 1 min on ice. Collect the contents of the tube by brief centrifugation and add the following in the order given:

Component	Volume (µl)
10X PCR buffer	2.5
25 mM $MgCl_2$	2.5
10 mM dNTP mix	1
0.1 M DTT	2.5
final volume	8.5

Note: For high GC content mRNA, use the Alternate Protocol given in section 6.1 to help reduce interference by secondary structure of mRNA.

The final volume of step 1 and 2 is 24 µl.

3. Mix gently, and collect the reaction by brief centrifugation. Incubate for 1 min at 42°C.
4. Add 1 µl of SuperScript™ II RT. Mix gently and incubate for 50 min at 42°C. **Note**: 30 min incubation is usually sufficient for short (<4 kb) mRNAs. Longer transcripts require at least 50 min to synthesize enough cDNA for a consistent signal in long PCR.

Note: Keep enzymes on ice during the procedure. Mix, and quickly centrifuge each component before use.

Final composition of the reaction:

20 mM Tris-HCl (pH 8.4)
50 mM KCl
2.5 mM $MgCl_2$
10 mM DTT
100 nM cDNA primer (GSP1)
400 µM each dATP, dCTP, dGTP, dTTP
1-5 µg (~40 ng/µl) RNA
200 units SuperScript™ II RT

Note: If you have >5 µg of total RNA, increase reaction volumes and amount of SuperScript™ II RT proportionately. If you have <1 µg of total RNA, no changes to the protocol are necessary. 50 to 500 ng of poly(A)$^+$ RNA may be substituted for total RNA in this protocol.

Seikagaku Corporation Exhibit 2004 - Page 120 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Note: Possible stopping points are noted in the protocol with the STOP icon.

5. Incubate at 70°C for 15 min to terminate the reaction.
6. Centrifuge 10 to 20 s and place the reaction at 37°C.
7. Add 1 µl of RNase mix, mix gently but thoroughly, and incubate for 30 min at 37°C.
8. Collect the reaction by brief centrifugation and place on ice.

The procedure may be stopped at this point and the reactions stored at -20°C. STOP

3.6 S.N.A.P. Column Purification of cDNA

Note: 1X wash buffer and 70% ethanol must be prepared prior to use and used cold (see section 3.4.3).

Note: The binding solution must be at room temperature for efficient binding of the DNA.

1. Equilibrate the binding solution to room temperature.
2. For each sample to be purified, equilibrate ~100 µl of sterilized, distilled water at 65°C for use in step 9.
3. Add 120 µl of binding solution (6 M NaI) to the first strand reaction.
4. Transfer the cDNA/NaI solution to a S.N.A.P. column. Centrifuge at 13,000 *x g* for 20 s.
5. Remove the cartridge insert from the tube and transfer the flowthrough to a microcentrifuge tube. Save the solution until recovery of the cDNA is confirmed. Place the cartridge insert back into the tube.
6. Add 0.4 ml of **COLD** (4°C) 1X wash buffer to the spin cartridge. Centrifuge at 13,000 x *g* for 20 s. Discard the flowthrough. Repeat this wash step **three** additional times.
7. Wash the cartridge two times with 400 µl of **COLD** (4°C) 70% ethanol as described in step 6.

Note: Failure to remove all the ethanol can result in poor recovery of the DNA.

Note: It is very important that the distilled water be at 65°C in order to maximize recovery of DNA.

8. After removing the final 70% ethanol wash from the tube, centrifuge at 13,000 x *g* for 1 min.
9. Transfer the spin cartridge insert into a fresh sample recovery tube. Add 50 µl of sterilized, distilled, water (**preheated to 65°C**) to the spin cartridge. Centrifuge at 13,000 x *g* for 20 s to elute the cDNA. STOP

3.7 TdT Tailing of cDNA

Variable amounts of purified cDNA from the S.N.A.P. column purification may be used in the TdT-tailing reaction. Factors, such as the amount of RNA used in the first strand reaction and relative abundance of the desired mRNA, should be considered. If desired, the cDNA pool may be concentrated by lyophilization and the entire contents used in the tailing reaction. To evaluate the specificity of the subsequent amplification reaction from the oligo-dC tail, inclusion of a control reaction that omits TdT is recommended.

1. Add the following components to each tube and mix gently:

Component	Volume (µl)
DEPC-treated water	6.5
5X tailing buffer	5.0
2 mM dCTP	2.5
S.N.A.P.-purified cDNA sample	10.0
final volume	24.0

Note: The 5X tailing buffer already contains $MgCl_2$. DO NOT add it separately.

2. Incubate for 2 to 3 min at 94°C. Chill 1 min on ice. Collect the contents of the tube by brief centrifugation and place on ice.

Note: The TdT has been optimized for 5' RACE. Increased amounts of TdT in the tailing reaction may inhibit PCR.

3. Add 1 µl TdT, mix gently, and incubate for 10 min at 37°C.

   Final composition of the reaction:

   10 mM Tris-HCl (pH 8.4)
   25 mM KCl
   1.5 mM $MgCl_2$
   200 µM dCTP
   cDNA
   TdT

Seikagaku Corporation Exhibit 2004 - Page 121 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

4. Heat inactivate the TdT for 10 min at 65°C. Collect the contents of the reaction by brief centrifugation and place on ice. STOP

3.8 PCR of dC-tailed cDNA

Tailed cDNA obtained from the preceding protocol may be amplified directly by PCR. Amplification of > 5 µl volumes of the tailing reaction requires appropriate adjustments for buffer, $MgCl_2$, and dNTP concentrations in the PCR.

1. Equilibrate the thermal cycler block to 94°C. In most cases, the "good start" procedure gives specific amplification products. For some target and primer sets, "hot start" has been reported to improve the specificity of the reaction (44,45).
2. Add the following to a 0.2 or 0.5-ml thin-wall PCR tube sitting on ice:

Note: It is very important to have the reaction mixture ice-cold to avoid nonspecific binding and extension of primers.

Component	Volume (µl)
sterilized, distilled water	31.5
10X PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl]	5.0
25 mM $MgCl_2$	3.0
10 mM dNTP mix	1.0
nested GSP2 (prepared as 10 µM solution)	2.0
Abridged Anchor Primer (10 µM)	2.0
dC-tailed cDNA	5.0
final volume	49.5

3. Add 0.5 µl of *Taq* DNA polymerase (5 units/µl) immediately before mixing.
4. Mix the contents of the tube (*Taq* DNA polymerase is added immediately before going into the thermal cycler) and overlay with 50 to 100 µl of mineral oil (if necessary).

   Final composition of the reaction:

   20 mM Tris-HCl (pH 8.4)
   50 mM KCl
   1.5 mM $MgCl_2$
   400 nM GSP2
   400 nM Abridged Anchor Primer
   200 µM each dATP, dCTP, dGTP, dTTP
   tailed cDNA
   2.5 units *Taq* DNA polymerase
5. Transfer tubes directly from ice to the thermal cycler preequilibrated to the initial denaturation temperature (94°C).
6. Perform 30 to 35 cycles of PCR. A typical cycling protocol for cDNA with ≤1 kb amplified region is:

PAD:	94°C for 1-2 min
Cycle:	
Denaturation:	94°C for 0.5-1 min
Annealing of primers:	55°C for 0.5-1 min
Primer extension:	72°C for 1-2 min
Followed by:	
Final extension:	72°C, 5-7 min
Indefinite hold:	5°C, until samples are removed.

7. Analyze 5-20 µl of 5' RACE products by agarose gel electrophoresis according to standard protocols, using appropriate size standards (19). Either TAE or TBE electrophoresis buffer may be used for the procedure. The volume of the sample used for analysis will depend on the volume and thickness of the sample well. If products will be extracted for reamplification, ultraviolet (UV) visualization of ethidium bromide-stained products should be performed using either a long wavelength (356-nm) UV or 302-nm wavelength source to minimize DNA nicking.

Note: The cycling protocol and annealing/extension temperatures may need to be optimized for each target sequence and/or thermal cycler. See section 2.2.7.

Seikagaku Corporation Exhibit 2004 - Page 122 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

3.9 Nested Amplification

Often a single PCR of 25 to 35 cycles will not generate enough specific product to be detectable by ethidium bromide staining. Increasing the number of cycles performed during the PCR beyond 35 cycles may generate numerous nonspecific products and is not recommended. Similarly, high target levels may exacerbate amplification of nonspecific products during PCR and contribute to the production of a heterologous smear of PCR products. Instead, a dilution of the original PCR (0.1 to 0.05%) can be re-amplified (7) using the AUAP or UAP and a nested GSP (see figures 1 and 2, see Chapter 4 for more information). If there is insufficient sequence information to design a nested GSP, re-amplification of gel purified, size-selected PCR products using the UAP, or AUAP, and original GSP is useful for enriching specific 5' RACE products or installation of dUMP-cloning sequences for UDG cloning. Nested PCR may also be conveniently conducted using a plug of agarose from the gel analysis of the initial 5' RACE reaction (see section 6.5) (63).

1. Dilute a 5 µl aliquot of the primary PCR into 495 µl TE buffer [10 mM Tris-HCl, (pH 8.0), 1 mM EDTA].
2. Equilibrate the thermal cycler block to 94°C.
3. Add the following to a 0.2 or 0.5-ml thin-wall PCR tube sitting on ice.

Component	Volume (µl)
sterilized, distilled water	33.5
10X PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl]	5.0
25 mM $MgCl_2$	3.0
10 mM dNTP mix	1.0
nested GSP (prepared as 10 µM solution)	1.0
AUAP or UAP (10 µM)	1.0
dilution of primary PCR product	5.0
final volume	49.5

Note: It is very important to have the reaction mixture ice cold to avoid nonspecific binding and extension of primers.

Note: UDG cloning requires amplification with the UAP and appropriately designed nested GSP. Use the AUAP if using an archaeobacterial DNA polymerase or long PCR enzyme mixture.

4. Add 0.5 µl of *Taq* DNA polymerase (5 units/µl) immediately before mixing.
5. Mix the contents of the tube (*Taq* DNA polymerase is added immediately before going into the thermal cycler) and overlay with 50 to 100 µl of mineral oil (if necessary).

   Final composition of the reaction:

   20 mM Tris-HCl (pH 8.4)
   50 mM KCl
   1.5 mM $MgCl_2$
   200 nM nested GSP
   200 nM UAP or AUAP
   200 µM each dATP, dCTP, dGTP, dTTP
   diluted primary PCR product
   2.5 units *Taq* DNA polymerase

6. Transfer tubes directly from ice to the thermal cycler preequilibrated to the initial denaturation temperature.
7. Perform 30 to 35 cycles of PCR.
8. Analyze 5 to 20 µl of the amplified sample, using agarose gel electrophoresis, ethidium bromide staining, and the appropriate molecular size standards.

Note: Cycle number will depend on the amount and complexity of the target. As few as 5 cycles can be used to install dUMP-sequences for UDG cloning.

Seikagaku Corporation Exhibit 2004 - Page 123 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Interpretation of Results

4

Analysis of 5' RACE Results: Following PCR, products may be analyzed by agarose gel electrophoresis (1% to 2%) and ethidium bromide staining. Band intensity and size distribution of resulting products depends on the specificity of GSPs used for cDNA synthesis and PCR, the complexity and relative abundance of target cDNA, and the PCR conditions used. Amplification products may vary from a single specific band to multiple discrete products to a broad diffuse smear. Incomplete cDNA synthesis, aberrant priming of GSPs during first strand synthesis or PCR, mispriming by the anchor primer, as well as primer-dimer and other PCR artifacts may contribute to the complexity of products obtained by 5' RACE. Identification of specific product bands may be complicated by the presence of nonspecific products that are dependent on both reverse transcription and dC-tailing (36). If sequences are available for use as internal probes, it is strongly recommended that Southern blot analysis be used to identify specific product bands. Specific products can also be identified using a diagnostic restriction endonuclease digestion if the amplified cDNA sequence contains a known restriction site.

5' RACE Controls: Several controls may facilitate interpretation of results. Products that result from amplification of contaminating genomic DNA can be identified from control reactions that omit RT. An alternative approach is to include control reactions that use genomic DNA as target (6). Specificity of the anchor primer for the oligo-dC tail should be examined by performing amplification reactions with cDNA subjected to dC-tailing both in the presence and absence of TdT. Additional controls that amplify dC-tailed cDNA using each primer individually (either the Abridged Anchor Primer or GSP2) may be useful in identifying nonspecific products that result from mispriming.

Nested Amplification: 5' RACE of rare messages may require additional PCR using a nested GSP and either the UAP or AUAP (see section 3.9). Generally, a dilution of the original PCR is used as target. A nested primer is composed of sequences located 3' to the original primer (GSP2). For 5' RACE, this would be an antisense primer that anneals closer to the mRNA 5'-end. Purification of the original PCR product from primers and primer-dimer products may significantly improve the specificity and efficiency of nested amplification procedures. Ultimately, the 5' RACE procedure should produce a single prominent band on an agarose gel. This may require additional rounds of PCR using successively nested GSPs.

Decisions regarding the design of a nested primer will depend on the amount of sequence information available for the target of interest and on the results of the original amplification reaction. When performing 5' RACE with a nested primer, sequences specific for downstream cloning manipulations (see section 3.4.2) must be designed into the nested GSP.

Seikagaku Corporation Exhibit 2004 - Page 124 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

5 Troubleshooting Guide

5.1 Testing the 5' RACE System Using the Control RNA and DNA

When using the 5' RACE System for the first time, we suggest performing an experiment using the control RNA to become familiar with the 5' RACE System procedure and to verify proper functioning of all components in the protocol, including your reagents and equipment for PCRs. The control RNA provided with the 5' RACE System is an 891-bp, *in vitro* transcribed RNA from the chloramphenicol acetyltransferase (CAT) gene that has been engineered to contain a 3' poly(A) tail. It may be used alone or added to your RNA preparation to test system performance in a background of heterologous nucleic acid. This is useful to test for the presence of contaminating nucleases. If desired, dilutions of the control RNA may be used to determine the sensitivity of the system or to model the abundance of the desired mRNA.

Note: Extreme care should be exercised when working with the control DNA so the 5' RACE reagents do not become contaminated. Use of aerosol barrier pipette tips significantly reduces the potential for the introduction of aerosols.

The control DNA was constructed by cloning the control 5' RACE product into pAMP1. Tailed cDNA was amplified using the 5' RACE Anchor Primer and control GSP2 containing additional UDG cloning sequences. The 4.8-kb pAMP1 5'RACE recombinant contains the oligo-dC tail sequence and may be used as a PCR positive control to verify the performance of the Abridged Anchor Primer, 5' RACE Anchor Primer, AUAP, UAP, or the control GSP3, in conjunction with the control GSP2. Alternately, it may be used to optimize PCR parameters with the Abridged Anchor Primer for your reaction conditions or thermal cycling device.

Two different PCRs are used to verify system performance. Conversion of first strand cDNA and recovery of cDNA after S.N.A.P. purification are assayed by a CAT cDNA-specific PCR using the control GSP3 and GSP2. Addition of the oligo-dC tail to purified control cDNA is assayed by PCR using the Abridged Anchor Primer and control GSP2. This strategy is depicted schematically in figure 6. Sequences for the control primers are presented in figure 4. The annealing sites for the control primers and resulting amplification products are shown in figure 5, panel 1. Note: The user may find it advantageous to adopt a similar RT-PCR strategy for their message and design an appropriate sense gene-specific primer (GSP3) to facilitate troubleshooting problems that may arise during 5' RACE with their message.

control GSP1	5'-TTG TAA TTC ATT AAG CAT TCT GCC-3'
control GSP2	5'-GAC ATG GAA GCC ATC ACA GAC-3'
control GSP3	5'-CGA CCG TTC AGC TGG ATA TTA C-3'

Figure 4. Sequences of the Control Primers.

Typical results for the procedure using the control RNA, both alone and in a background of 1 µg HeLa total RNA, are shown in figure 5, panel 2. A distinct 711-bp 5' RACE PCR product (solid arrow) should be visible by ethidium bromide staining. Other products, generally visible as faint bands or a diffuse smear, can result from spurious priming by GSP2 or the anchor primer, incomplete cDNA synthesis, and primer-dimer artifacts. If the control RNA is used in a background of heterologous RNA, nonspecific 5' RACE products, that are dependent on both RT and TdT, may be observed. This effect is illustrated by the 310-bp HeLa-derived 5' RACE product (open arrow). The presence of these nonspecific but genuine 5' RACE products is primarily a function of the specificity of the GSPs and emphasizes the need for characterization and enrichment of specific amplification products prior to cloning or sequencing.

Seikagaku Corporation Exhibit 2004 - Page 125 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

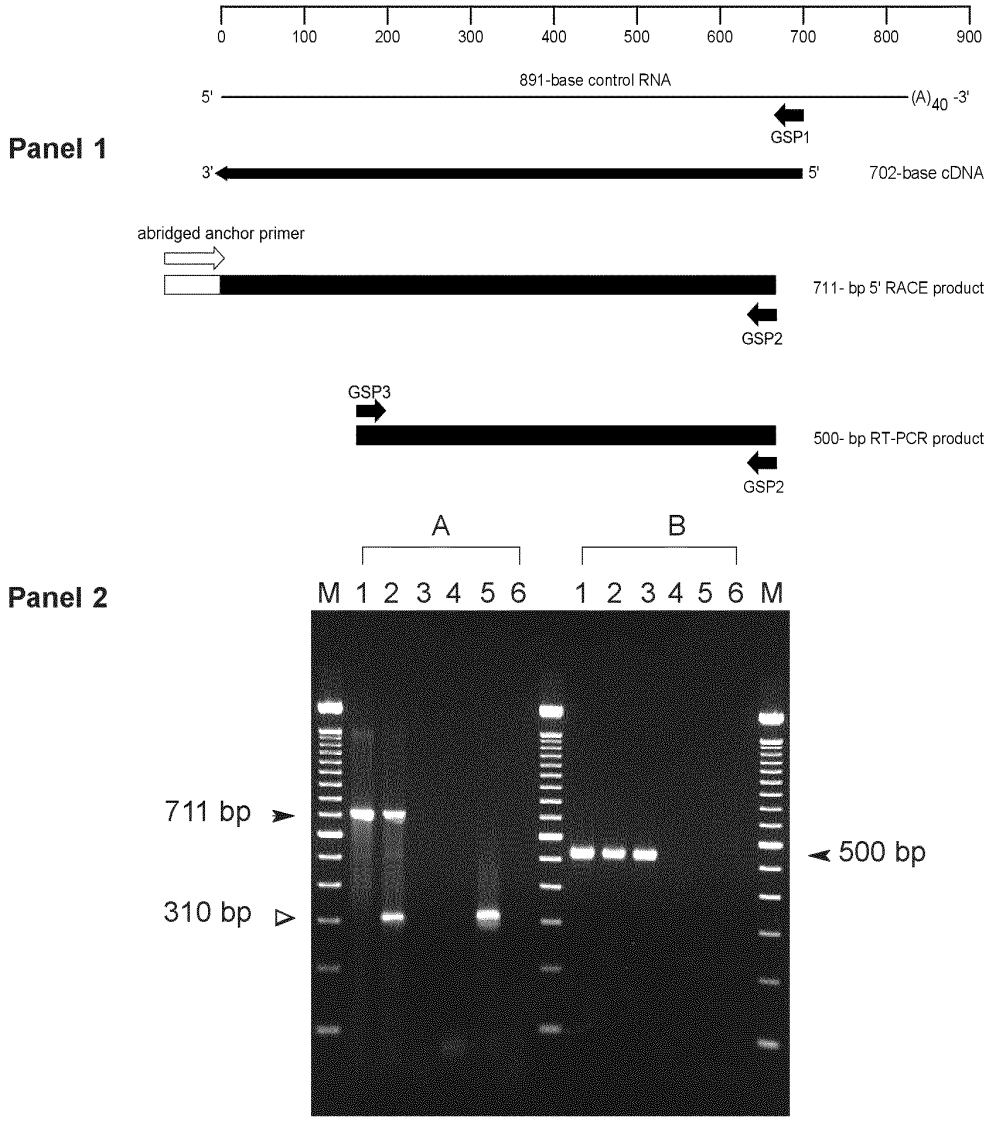


Figure 5. Agarose Gel Analysis of 5' RACE Products Using Control RNA. Panel 1. Schematic of the CAT RNA control system. Arrows designate 3' termini. The oligo-dC anchor is depicted as an open box. Panel 2. Control RNA (1 x 10^7 copies) was added to 1-µg aliquots of total RNA isolated from HeLa cells and subjected to 5' RACE amplified through 35 cycles with either (A) the anchor primer and control GSP2 or (B) control GSP2 and control GSP3 as described in this section. Reaction products were analyzed by gel electrophoresis in a 1.5% agarose gel (1X TBE) and stained with 0.5 µg/ml ethidium bromide. Lane 1, 1 x 10^7 copies of control RNA. Lanes 2 to 5 contain 1 µg of HeLa total RNA. Lanes 2-4 contain 1 x 10^7 copies of control RNA. Lane 3, no TdT control. Lane 4, no RT control. Lane 5, HeLa RNA only. Lane 6, water control. Target level is expressed as input values of RNA. Analyzed products represent 0.5% of target input given above. Lane M, GIBCO BRL 100 bp DNA Ladder. Open arrow, HeLa-derived 5' RACE product.

Seikagaku Corporation Exhibit 2004 - Page 126 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

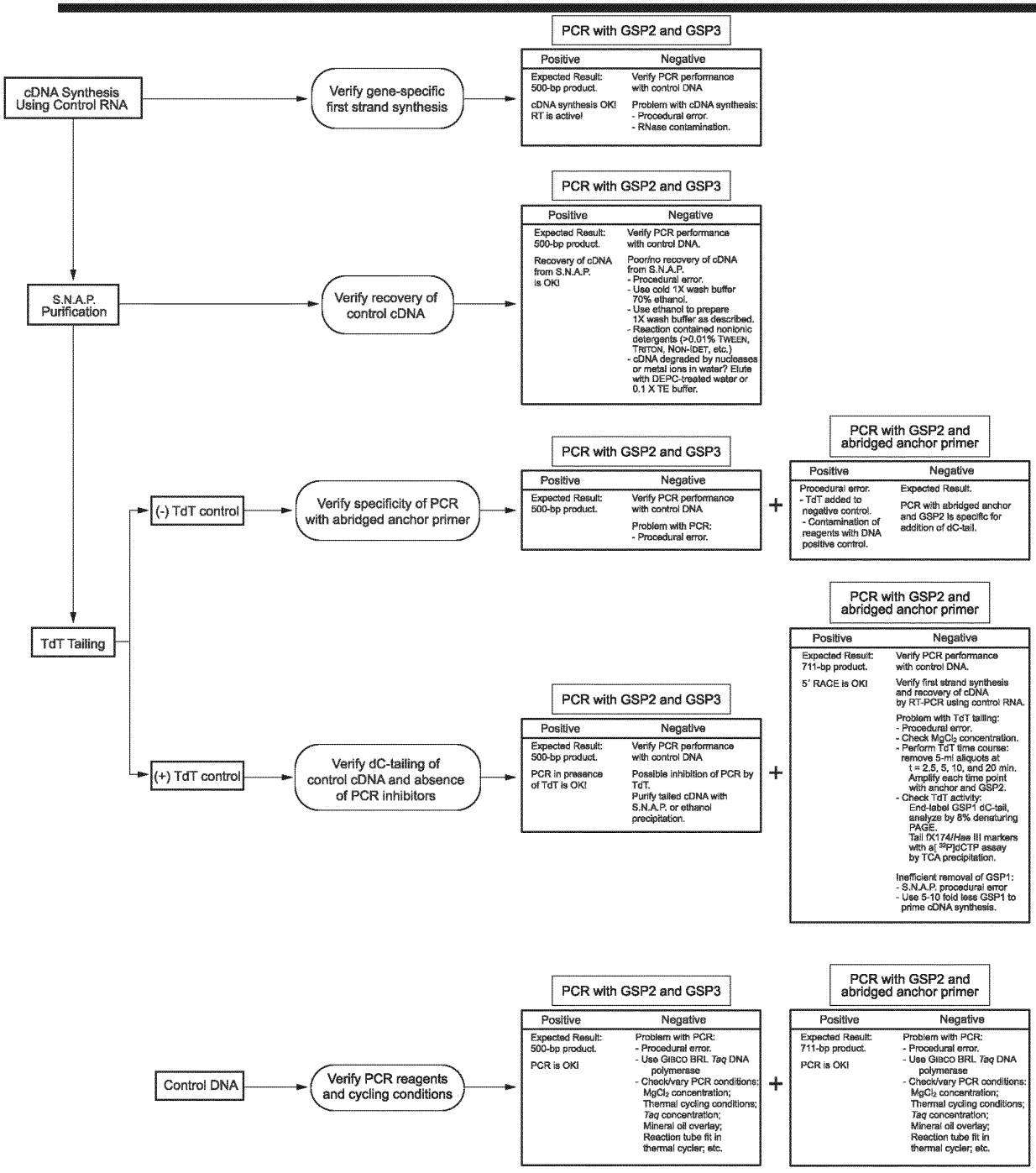


Figure 6: Flow Diagram for Troubleshooting the 5' RACE System Using the Control RNA and DNA.

Seikagaku Corporation Exhibit 2004 - Page 127 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

5.2 Control First Strand cDNA Synthesis

The RNA template for this step is control RNA and the primer is control GSP 1. If you wish to test performance in a background of your RNA we suggest doing two first strand reactions: (Control RNA + Control GSP 1) and (Control RNA +Control GSP 1 in the presence of your sample RNA). Adjust the volume of DEPC-treated water appropriately so that the final volume in step 1 is still 15.5 µl.

Note: Mix and quickly centrifuge each component before use.

1. Add the following to a 0.5-ml microcentrifuge tube:

Component	Volume (µl)
Control GSP1 (1 µM)	2.5
control RNA	1.0
DEPC-treated water	12.0
final volume	15.5

2. Incubate the mixture 10 min at 70°C to denature RNA. Chill 1 min on ice. Collect the contents of the tube by brief centrifugation and add the following in the order given:

Component	Volume (µl)
10X PCR buffer	2.5
25 mM $MgCl_2$	2.5
10 mM dNTP mix	1.0
0.1 M DTT	2.5
final volume	8.5

   The combined total volume of steps 1 and 2 is 24 µl.

3. Mix gently, and collect the reaction by brief centrifugation. Incubate for 1 min at 42°C.
4. Add 1 µl of SuperScript™ II RT. Mix gently and incubate for 50 min at 42°C.
5. Incubate at 70°C for 15 min to terminate the reaction.
6. Centrifuge 10 to 20 s and place the reaction at 37°C.
7. Add 1 µl of RNase Mix, mix gently, and incubate for 30 min at 37°C.
8. Collect the reaction by brief centrifugation and place on ice.
9. Transfer a 2-µl aliquot of the control first strand cDNA to a 1.5-ml microcentrifuge tube containing 998 µl of TE buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA]. Label the tube "A" and retain for later PCR.
10. Purify the remainder of the control DNA using the S.N.A.P. column. STOP

5.3 S.N.A.P. Column Purification of the Control cDNA

1. Equilibrate the binding solution to room temperature.

Note: The binding solution must be at room temperature for efficient binding of the DNA.

2. Equilibrate ~100 µl of distilled water at 65°C (for each sample to be purified) for use in step 9.
3. Add 120 µl of binding solution (6 M NaI) to the remainder of the control first strand reaction.
4. Transfer the cDNA/NaI solution to a S.N.A.P. column. Centrifuge at 13,000 x *g* for 20 s.
5. Remove the cartridge insert from the tube and transfer the flowthrough to a microcentrifuge tube. Save the solution until recovery of the cDNA is confirmed. Place the cartridge insert back into the tube.
6. Add 0.4 ml of **COLD** (4°C) 1X wash buffer to the spin cartridge. Centrifuge at 13,000 x *g* for 20 s. Discard the flowthrough. Repeat this wash step **three** additional times.

Note: 1X wash buffer and 70% ethanol must be prepared prior to use and used cold (see section 3.4.3 and 3.4.4).

7. Wash the cartridge two times with 400 µl of **COLD** (4°C) 70% ethanol as described in step 6.

Seikagaku Corporation Exhibit 2004 - Page 128 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Troubleshooting Guide

Note: Failure to remove all the ethanol can result in poor recovery of the DNA.

Note: It is very important that the distilled water be at 65°C in order to maximize recovery.

Note: The 5X tailing buffer has $MgCl_2$ already present. DO NOT add it separately.

Note: Extreme care should be exercised when working with the control DNA so the 5' RACE reagents do not become contaminated. Use of aerosol barrier pipet tips significantly reduces the potential for the introduction of aerosols.

8. After removing the final 70% ethanol wash from the tube, centrifuge at 13,000 x *g* for 1 min.
9. Transfer the spin cartridge insert into a fresh sample recovery tube. Add 50 µl of distilled water (**preheated to 65°C**) to the spin cartridge. Centrifuge at 13,000 x *g* for 20 s to elute the cDNA.
10. Transfer a 5-µl aliquot of the purified control first strand cDNA to a 0.5-ml microcentrifuge tube containing 495 µl of TE buffer. Label the tube "B" and retain for later PCR. STOP

5.4 TdT Tailing of the Control First Strand cDNA

A control reaction that omits TdT is included because amplification of cDNA from this control reaction can provide important data for troubleshooting RACE. For example, inclusion of a control reaction that omits TdT will help evaluate the specificity of the amplification reaction from the oligo-dC tail. Label two 0.5-ml microcentrifuge tubes "C" and "D", respectively.

1. In order to more sensitively test the efficiency of the tailing reaction, prepare a 100-fold dilution of the purified control cDNA. Add 1 µl of the cDNA to a 0.5-ml tube containing 99 µl of sterilized, distilled water.
2. Add the following components to each tube:

Component	Volume (µl)
DEPC-treated water (or sterile, distilled water)	6.5
5X Tailing Buffer	5.0
2 mM dCTP	2.5
S.N.A.P.-purified control cDNA (1:100 dilution)	10.0
final volume	24

3. Incubate for 2 to 3 min at 94°C. Chill 1 min on ice. Collect the contents of the tube by brief centrifugation and place on ice.
4. Add 1 µl TdT to tube(s) "C".
5. Add 1 µl DEPC-treated water to tube(s) "D".
6. Gently mix the contents of each reaction and incubate for 10 min at 37°C.
7. Heat inactivate the TdT for 10 min at 65°C. Collect the contents of the reaction by brief centrifugation and place on ice. Retain each reaction for later PCR. STOP

5.5 PCR of cDNA, Tailed cDNA and Control DNA

The PCR of the various aliquots from each step and the tailed cDNA product(s) is done using two sets of primers in two separate PCR mixes. An additional reaction that contains 5 µl of the control DNA as target is used as a positive control to verify PCR. One set of primers will be control GSP2 and control GSP3. Specific amplification of the control first strand cDNA product, or control DNA, results in a prominent 500-bp band when analyzing products by agarose gel electrophoresis and ethidium bromide staining. The other set of primers will be control GSP2 and Abridged Anchor Primer and should result in a prominent 711-bp product (and no product from the no TdT control) for both the tailed cDNA and the control DNA. The sample templates to be amplified are:

"A" cDNA from the first strand reaction (checks the efficiency of the cDNA synthesis reaction)

"B" S.N.A.P. eluate sample: (checks the recovery of cDNA from S.N.A.P. column)

"C" Tailed cDNA: (actual 5' RACE product; checks the efficiency of tailing)

"D" No TdT control: (checks the specificity of amplification of the tailed product, and presence of inhibitors of PCR)

Control DNA:(checks PCR with both primer sets)

You may have more than one tube for each set, depending on the number of first

Seikagaku Corporation Exhibit 2004 - Page 129 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

strand reactions you have.

1. Make two PCR mixes, one for each primer set:
   Mix I = Control GSP2 and GSP3
   Mix II = Control GSP2 and Abridged Anchor Primer

Add the following components on ice to an appropriately sized sterile tube, *e.g.* 1.5-ml microcentrifuge tube. Make enough mix for n+2 reactions where n = number of template samples to be amplified. (Minimum value for n is 5: A + B +C + D + Control DNA). Volumes given in the table are for 50 µl PCRs.

Mix I (GSP2 and GSP3)

Component	Volume (µl) per reaction	Volume (µl) 7X mix n=5	Volume (µl) 12X mix n=10
DEPC-treated water	33.5	234.5	402.0
10X reaction buffer	5.0	35.0	60.0
25 mM $MgCl_2$	3.0	21.0	36.0
10 mM dNTP mix	1.0	7.0.	12.0
control GSP2 (10 µM)	1.0	7.0	12.0
control GSP3 (10 µM)	1.0	7.0	12.0
Taq DNA Polymerase (5 units/µl)	0.5	3.5	6.0
final volume	45.0	315.0	540.0

Note: Add the *Taq* DNA polymerase just before you are ready to pipet the mix into the PCR tubes containing template.

Mix II (GSP2 and Abridged Anchor Primer)

Component	Volume (µl) per reaction	Volume (µl) 7X mix n=5	Volume (µl) 12X mix n=10
DEPC-treated water	31.5	220.5	378.0
10X reaction buffer	5.0	35.0	60.0
25 mM $MgCl_2$	3.0	21.0	36.0
10 mM dNTP mix	1.0	7.0	12.0
control GSP2 (10 µM)	2.0	14.0	24.0
Abridged Anchor Primer	2.0	14.0	24.0
Taq DNA Polymerase 5 units/(µl)	0.5	3.5	6.0
final volume	45.0	315.0	540.0

2. Pipet 5 µl of template samples (A,B,C,D, and control DNA) into two sets appropriately labeled thin-walled 0.2 or 0.5 ml PCR tubes. There will be one set for each PCR mix.
3. Add 45 µl of Mix I to one set of tubes and add 45 µl of Mix II to the second set of tubes.
4. Mix the contents of the tubes and overlay with 50 to 100 µl of mineral oil (if necessary).
5. Equilibrate the thermal cycler block to 94°C.
6. Transfer tubes directly from ice to the hot thermal cycler.
7. Perform 35 cycles of PCR:
   Denature 94°C for 1 min
   Anneal 63°C for 30 s
   Extend 72°C for 2 min

Note: Temperatures may vary slightly among thermal cyclers. The annealing temperature can be adjusted (61° - 65°C) to compensate for this variation.

8. Incubate the reaction for 10 min at 72°C following the last cycle of PCR, then maintain reactions at 4°C.
9. Analyze 5-20 µl of 5' RACE products by agarose gel electrophoresis according to standard protocols, using appropriate size standards (19). Either TAE or TBE electrophoresis buffer may be used for the procedure. The volume of the sample used for analysis will dependent on the volume and thickness of the sample well. If products will be extracted for reamplification, ultraviolet (UV) visualization of ethidium bromide-stained products should be performed using either a long wavelength (356-nm) UV or 302-nm wavelength source to minimize DNA nicking.

Seikagaku Corporation Exhibit 2004 - Page 130 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Troubleshooting Guide

5.6. General Troubleshooting Guidelines for the 5' RACE System

Problem	Possible Cause	Suggested Remedy
No bands after electrophoretic analysis of amplified products	5' RACE amplification product may be present but in too low a concentration for detection by ethidium bromide staining	Perform Southern blot analysis of amplifiction products using internal sequence as probe.
		Amplify agarose gel purified material of expected size range using the AUAP or UAP, and user's GSP2.
		Perform nested primer amplification from either purified amplification products, dilution of original PCR (0.05%), or agarose gel plug using the UAP,or AUAP, and user's nested GSP.
		For further discussion, see Chapter 4, Interpretation of Results and section 6.5. Refer to Chapter 6 for specific protocols.
	Procedural error in first strand cDNA synthesis, purification of cDNA product, TdT tailing, or PCR	Use the control RNA to verify conversion of first strand product, recovery following purification and dC tailing.
		Design sense GSP from available 5'-mRNA sequence data. Verify first strand conversion of desired message by PCR using two gene-specific primers.
	Inhibitors of RT present	Remove inhibitors by ethanol precipitation of the mRNA preparation before the first strand reaction. Include a 70% (v/v) ethanol wash of the mRNA pellet.
		Note: Inhibitors of RT include sodium dodecyl sulfate (SDS), EDTA, guanidinium salts, and glycerol. Inhibitors of M-MLV-RT include sodium pyrophosphate and spermidine. SuperScript™ RT is inhibited 50% by 0.0025% SDS, 1 mM EDTA , 15 mM guanidine isothiocyanate, 17% DMSO, 50% glycerol, 5% formamide, 4 µg/ml heparin and 4 mg/ml glycogen (52).
		Test for the presence of inhibitors by mixing 1 µg of control RNA with 1 µg of sample RNA and comparing yields of first strand cDNA or by PCR of control band.
	Target mRNA has secondary structure that interferes with annealing of GSP1	Redesign GSP1 and/or use section 6.1.1.
	Target mRNA contains strong transcriptional pauses	Maintain an elevated temperature after the annealing step and increase the temperature of first strand reaction (up to 50°C). See section 6.1.1.

Seikagaku Corporation Exhibit 2004 - Page 131 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Problem	Possible Cause	Suggested Remedy
No bands after electrophoretic analysis of amplified products	RNase contamination	Perform first strand synthesis with the control RNA both alone and added to sample RNA to determine if RNase is present. Assay for control cDNA by PCR using control GSP2 and GSP3. See section 5.1.
		Maintain aseptic conditions to prevent RNase contamination. See section 5.7.
		Use placental RNase inhibitor during first strand cDNA synthesis.
		Note: Placental RNase inhibitor requires sulfhydryl reagents for maximal RNase binding activity (19). Always add RNase inhibitor to reactions after the addition of dithiothreitol. Treatments which denature the protein, such as high temperature incubation, exposure to oxidizing conditions, or repeated freezing and thawing, can release RNases initially bound by the inhibitor. These RNases may subsequently degrade RNA preparations in downstream procedures.
	Inefficient tailing of cDNA	Perform TdT time course: remove 5-µl aliquots at t = 2.5, 5, 10, and 20 min. Amplify each time point using abridged anchor primer and GSP2.
	cDNA did not tail due to strong secondary structure of 3'-end	Denature cDNA prior to tailing by incubation at 94°C for 2 to 3 min, then quick chill on ice. Perform the tailing reaction on ice for 1 h to keep 3'-end denatured and available for TdT.
		Alternatively, use of cosolvents *e.g.* DMSO (≤ 20%) has been been found to increase efficiency of tailing.
	Inhibition of PCR by TdT	Purify tailed cDNA by ethanol precipitation in the presence of inert carrier, *i.e.* glycogen, and 2.5 M NH_4OAc.
	Polysaccharides and small RNAs coprecipitate with mRNA	Ethanol precipitate the RNA preparation; treat the pellet as described in section 6.4.
	Polymerase used in PCR was from an archaeobacterium and dUMP primers were used	Use abridged anchor primer, or AUAP, and non-dUMP-containing GSP in PCR using eLONGase® reagents or an archaeobacterial polymerase.
		Use *Taq* DNA polymerase for PCR with the UAP and dUMP-containing primers.

Seikagaku Corporation Exhibit 2004 - Page 132 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Troubleshooting Guide

Problem	Possible Cause	Suggested Remedy
Agarose gel analysis shows strong primer-dimer product, but no visible gene-specific product	Inefficient removal of GSP1 during S.N.A.P. procedure	Design GSP1 according to the guidelines discussed in section 3.4.2. Reduce the concentration of GSP1 used in first strand synthesis. Remove dC-tailed GSP1 with an additional S.N.A.P. purification after TdT tailing and second strand synthesis as described in section 6.2. **Note:** dCTP and the 5' RACE anchor primer may be substituted for dATP and the 3' RACE system adapter primer described in this protocol. Substitute dU for dT in GSP1. Degrade dU-GSP1 following cDNA synthesis with UDG. Purify cDNA by ethanol precipitation in the presence of inert carrier, *i.e.* glycogen, and NH_4OAc. **Note:** This technique may be useful for 5' RACE of short cDNAs that bind poorly to the S.N.A.P. column.
	Primer-dimer product between the anchor primer and GSP2	Minimize nonspecific annealing of primers by initiation of PCR at an elevated temperature (75°C to 80°C). See section 3.9.
Unexpected bands after electrophoretic analysis of "nested" amplification products	Contamination by genomic DNA Spurious priming in the PCR	Pretreat RNA as described in section 6.3. Vary the parameters of the PCR: increase annealing temperature, decrease annealing time, reduce $MgCl_2$ concentration, etc. (24-26, 38, 42, 43). Minimize nonspecific annealing of primers by initiation of PCR at an elevated temperature (75°C to 80°C). See section 3.9. Perform PCR control amplifications as discussed in Chapter 4.
Gene-specific 5' RACE products appear as a smear. Unable to isolate full length 5' RACE product	RNA preparation is degraded or of poor quality	Isolate new preparation of total RNA.
Absence of discrete product bands. 5' RACE products appear as a heterogeneous smear	Potentially, a normal result	Identfiy gene-specific products by Southern blot hybridization. Enrich for specific products by reamplification of gel-purified material or nested primer amplification.

Seikagaku Corporation Exhibit 2004 - Page 133 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Problem	Possible Cause	Suggested Remedy
Absence of discrete product bands. 5' RACE products appear as a heterogeneous smear	High levels of poly $(A)^+$ RNA (≥1 µg) used for first strand synthesis may contribute to nonspecific products. RNA-primed, (GSP1 independent) cDNA may compete for TdT and also for anchor primer during the PCR.	Reduce target level or optimize tailing conditions. Perform TdT time course: remove 5-µl aliquots at t = 2.5, 5, 10, and 20 min. Amplify each time point using abridged anchor primer and GSP2.
5' RACE product does not correspond to known mRNA sequence. Product is dependent on TdT addition of dC-tail	Target mRNA contains strong transcriptional pauses	Maintain an elevated temperature after the annealing step and increase the temperature of first strand reaction (up to 50°C). See section 6.1.1. Use cosolvents *e.g.*, 5-10% DMSO or 10-20% glycerol compatible with the RT in first strand reaction (52) to help eliminate secondary structure while maintaining RT activity.
5' RACE product does not correspond to full length mRNA sequence. Product is not dependent on TdT addition of dC-tail	Internal mispriming by the anchor primer at C-rich sequence	Deoxyinosine-containing anchor primer may not be suitable for amplification of C-rich cDNA. Use alternate 5' RACE strategy. See section 6.2.
5' RACE product does not correspond to known mRNA sequence.	Aberrant priming of gene-specific primers	Perform control amplification reactions as discussed in Chapter 4, 5' RACE Controls. Verify identity of 5' RACE product by hybridization or diagnostic restriction endonuclease cleavage prior to cloning or sequencing. Increase the specificity of GSP1 by increasing the temperature of first strand synthesis (up to 50°C). Optimize PCR conditions to maximize specificity of GSP2 and anchor primer. Increase annealing temperature, decrease annealing time, decrease $MgCl_2$ concentration, *etc.* Design new GSP1 and/or GSP2. See section 3.4.2.

Seikagaku Corporation Exhibit 2004 - Page 134 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Troubleshooting Guide

Problem	Possible Cause	Suggested Remedy
Poor cloning efficiency	Inefficient ligation	Increase the incubation time in the ligation reaction; decrease the temperature.
		Remove dNTPs prior to ligation.
		Use a ligation-free cloning method such as the CloneAmp system.
	Poor restriction endonuclease digestion due to residual bound *Taq* DNA Polymerase	Treat PCR products with proteinase K (64). See section 6.6.
	Fill-in of overhangs by residual *Taq* DNA polymerase	Extract PCR products with phenol:chloroform and purify by ethanol precipitation or treat with proteinase K (see above) before restriction endonuclease digestion.
	Clone is unstable in host cells	Try different strain of bacterial cells such as Stbl2™ competent cells.

Seikagaku Corporation Exhibit 2004 - Page 135 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

5.7 Minimizing RNase Contamination

Successful cDNA synthesis demands an RNase-free environment at all times, which will generally require the same level of care used to maintain aseptic conditions when working with microorganisms. Several additional guidelines should be followed:

1. Never assume that anything is RNase-free, except sterilized pipettes, centrifuge tubes, culture tubes, or any similar labware that is explicitly stated to be sterile. Wear latex gloves for all manipulations involving RNA.
2. Dedicate a separate set of automatic pipettes for manipulating RNA and the buffers and enzymes used to synthesize cDNA.
3. Avoid using any recycled glassware unless it has been specifically rendered RNase-free by rinsing with 0.5 N NaOH followed by copious amounts of sterilized, distilled water. Alternatively, bake glassware at 150°C for 4 h.
4. Microcentrifuge tubes can generally be taken from an unopened box, autoclaved, and used for all cDNA work. RNase-free microcentrifuge tubes can be purchased from several suppliers. If it is necessary to decontaminate untreated tubes, soak the tubes overnight in a 0.01% (v/v) aqueous solution of diethylpyrocarbonate (DEPC), rinse them with sterilized, distilled water, and autoclave them.
5. If made with RNase-free labware, most solutions can be made from reagent-grade materials and distilled water, and then autoclaved. Prepare heat-sensitive solutions using sterilized, distilled water, and filter them to 0.2 µm using sterilized, disposable filterware.

Note: Buffers containing primary amines (such as Tris) cannot be effectively treated with DEPC.

6. If all else fails, most aqueous buffer solutions can be treated with 0.01% (v/v) DEPC and autoclaved.
7. Use aerosol-resistant pipet tips.

5.8 T_m Values for 5' RACE and Control Primers

Oligonucleotide primer T_ms can vary widely depending on the method of calculation. A comparison of T_m determinations (°C) for the 5' RACE Anchor Primer, Abridged Anchor Primer, UAP, AUAP, control GSP2, and control GSP3 using the 2(AT) + 4(GC), %GC, and nearest neighbor methods are summarized in table 1. See figure 2 for sequences of the 5' RACE primers, and figure 4 for sequences of the control primers. The nearest neighbor analysis method (65) relies not only on the quantity of each base, but on the primary sequence of the DNA. Comparison of experimentally derived T_ms to that predicted by the available estimates for a number of oligonucleotides has found the nearest neighbor method to be the most accurate (22,66). However, the precise reaction conditions should be considered in the calculation. Other methods are either limited by the size of the oligonucleotide for which they are valid or derived for salt and temperature conditions that are different from those for PCR (see legend, table 1, and 19,67-70).

Seikagaku Corporation Exhibit 2004 - Page 136 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Troubleshooting Guide

Table 1. Comparison of T_m Values for 5' RACE and Control Primers calculated by different methods

Primer	2(AT) + 4(GC)[1]	%GC[2]	NN[3]	NN $_{PCR}$[4]
GI Anchor Region	52	72	60.5	66.6
Anchor Primer	150	90.3	93.9	90.1
Abridged Anchor Primer	118	90.0	94.4	92.5
UAP	98	83.8	75.8	75.0
AUAP	66	79.3	67.8	71.5
GSP2	64	75.3	64.5	69.0
GSP3	66	75.4	66.5	69.2

T_m values (°C) for 5' RACE and control primers were calculated using OLIGO™ 5.0 according to the formulae given below. The GI Anchor Region comprises the 16-base deoxyinosine-containing sequence at the 3'-end of the Abridged Anchor or Anchor Primer. The efficiency and specificity of annealing of this sequence to the dC-tail in the early PCR cycles is critical to the overall efficiency of 5' RACE. The disparity of T_m values obtained using different formulae highlights the importance of empirical determination of optimal annealing temperature for any given primer pair.

1) The T_m of many oligonucleotides can be estimated by assigning 2°C to each A and T and 4°C to each G and C (63,64).

   2(AT) + 4(GC): 2 x (numbers of A's + number of T's) + 4 x (number of C's + number of G's).

   This estimate was derived for hybridizations in high salt concentration (1M) and is only valid for oligonucleotides less than 18 bases in length (35).

2) Another method for estimating the T_m is based on the percent GC in the oligonucleotide (35,65,66). The formula was developed for hybridization to DNA immobilized on a solid support using molecules under 100 bases in length in conditions of high salt.

   %GC: Tm = 81.5 + 16.6 log[salt$^+$] + 0.41[%GC] -(675 / number of bases in oligonucleotide)

3) NN: nearest neighbor T_m (19) calculated using second order hybridization kinetics, 100 pM primer 1 M Na^+.

4) NN $_{PCR}$: nearest neighbor T_m (19) calculated for standard PCR conditions using first order hybridization kinetics, 200 nM primer, 50 mM Na^+, 0.7 mM free Mg^{++} (effective [Na^+] = 155.6 mM).

Seikagaku Corporation Exhibit 2004 - Page 137 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Additional Protocols

6

6.1 5' RACE of G:C Rich cDNA

The 5' RACE System protocol provides a rapid and reliable solution for capturing 5'-end mRNA sequences. The physical properties of each message, however, will influence the efficacy of the 5' RACE procedure. Some messages may require alternative strategies. Poly(dC) tracts which may be present in G:C rich cDNA can serve as effective priming sites for the deoxyinosine-containing anchor primer. PCR of such a template results in truncated products. In this case, a modification of the dATP-tailing method described by Frohman *et al.* (1, 6) is recommended.

Classical 5' RACE methodologies use the PCR process to install adapter priming sequences and amplify the cDNA from the homopolymer tail. However, PCR from homopolymer sequences, and especially oligo(dT), can be highly problematic. Because of these difficulties, a second-strand synthesis procedure is used to install the adapter priming site prior to PCR. An oligo(dT)-containing primer such as the 3' RACE Adapter Primer (AP) is used to prime second-strand synthesis (figure 7). Following purification, 5'-end sequences are amplified using the AUAP, or UAP, and a user-defined, nested gene-specific primer. This single-cycle installation, followed by a purification step to remove excess oligo(dT) primer, removes the potential for internal mispriming by the primer during each PCR cycle, and results in a higher proportion of "full-length" amplification products.

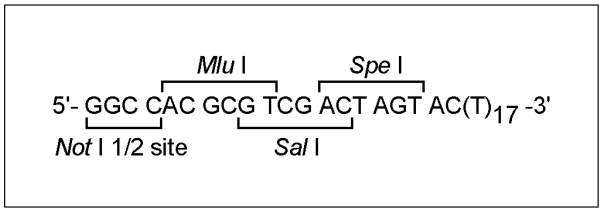


Figure 7. Sequence of the 3' RACE Adapter Primer.

6.1.1 Alternative Protocol for First Strand cDNA Synthesis of Transcripts with High GC Content

Problems with 5' RACE due to secondary structure of the target mRNA often can be overcome by increasing the volume and temperature of the RT reaction. To avoid secondary RNA structure, shift the primer/RNA mix directly from 70°C to 50°C and prewarm the complete 2X reaction mix to 50°C before adding it to the primer and RNA. Use of a thermal cycler simplifies the multiple temperature shifts in RT-PCR and can help to prevent formation of secondary structure in RNA.

First Strand Synthesis Conditions

20 mM Tris-HCl (pH 8.4 at 22°C)
50 mM KCl
2.5 mM $MgCl_2$
10 mM DTT
0.5 mM each dATP, dCTP, dGTP, and dTTP
100 nM GSP1
1-5 µg total RNA
200 units SuperScript™ II RT
Reaction volume: 50 µl

Seikagaku Corporation Exhibit 2004 - Page 138 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

1. Mix and briefly centrifuge each component before use.
2. Combine 1-5 µg of total or 50-500 ng of poly(A)$^+$ - RNA and DEPC-treated water to a final volume of 20 µl in a 0.2 or 0.5-ml microcentrifuge tube (or thin-walled PCR tube if using a thermal cycler).
3. Add 5 µl of the 1 µM GSP1 solution, mix gently, and collect by brief centrifugation.
4. Heat each sample to 70°C for 10 min and immediately transfer to 50°C.
5. In a separate tube, assemble the following reaction mixture adding each component in the indicated order. For n samples make n+1 reaction mix. The table below shows volumes for n=1.

Component	Volume (µl) 1 reaction	Volume (µl) n +1
DEPC-treated water	6.5	13.0
10X PCR buffer	5.0	10.0
25 mM $MgCl_2$	5.0	10.0
10 mM dNTP mix	2.5	5.0
0.1 M DTT	5.0	10.0
final volume	24.0	48.0

6. Prewarm the reaction mix to 50°C immediately before use.
7. Add 24 µl of the prewarmed reaction mixture and 1µl (200 units) of SuperScript™ II to the RNA/primer mixture to each sample tube, mix gently and continue to incubate at 50°C for 50 min.
8. Terminate the reactions by incubating the tubes at 70°C for 15 min. Chill on ice.
9. Collect the reactions by brief centrifugation. Add 1 µl of RNase mix to each tube and incubate for 30 min at 37°C.
10. Add 225 µl of binding solution (6 M NaI) and proceed to section 3.6.

6.2 dA-Tailing of cDNA

Variable amounts of purified cDNA may be used in the TdT-tailing reaction depending on the application. Factors such as the amount of RNA used in the first strand reaction and relative abundance of the desired mRNA should be considered. If desired, the cDNA pool may be concentrated by lyophilization and the entire contents used in the tailing reaction. Inclusion of a control reaction that omits TdT is recommended.

1. Add the following to a 0.5-ml microcentrifuge tube on ice:

Component	Volume (µl)
DEPC-treated water	6.5
5X Tailing Buffer	5.0
2 mM dATP	2.5
S.N.A.P.-purified cDNA sample	10.0
final volume	24.0

2. Incubate at 94°C for 2 to 3 min. Chill for 1 min on ice. Collect the contents of the tube by brief centrifugation and place on ice.
3. Add 1 µl of TdT, mix gently, and incubate at 37°C for 10 min.
   Final composition of the reaction:
   10 mM Tris-HCl (pH 8.4)
   25 mM KCl
   1.5 mM $MgCl_2$
   200 µM dATP
   TdT
4. Heat-inactivate the TdT for 10 min at 65°C. Collect the contents of the reaction by brief centrifugation and place on ice.

Seikagaku Corporation Exhibit 2004 - Page 139 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

6.2.1 Second-Strand Synthesis

This protocol has been adapted for enzymes supplied with the 5' RACE system. In general, SuperScript™ will utilize a DNA template with equivalent efficacy to that of an RNA template. However, the user may want to consider the use of other DNA-directed DNA polymerases and appropriately adjust reaction conditions (49).

1. Add the following components to the TdT reaction (25 µl):

Component	Volume (µl)
sterile, distilled water	8.0
10X PCR buffer[200 mM Tris-HCl (pH 8.4), 500 mM KCl]	4.0
25 mM $MgCl_2$	3.5
0.1 M DTT	5.0
10 mM dNTP mix	2.5
3' RACE AP (10 µM)	1.0
final volume	24.0

2. Equilibrate the reaction at 50°C for 2 min.
3. Add 1 µl of SuperScript™ II RT.
4. Incubate the reaction for 50 min at 50°C.
5. Add 225 µl of binding solution (6 M NaI) to the second strand reaction. Purify second-strand product using a S.N.A.P. column as described in section 3.6. Elute second-strand product using warm (65°C) TE buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA].

6.2.2 PCR of cDNA

1. Equilibrate the thermal cycler block to 94°C. In most cases, the "good start" procedure gives specific amplification products. For some target and primer sets, "hot start" has been reported to improve the specificity of the reaction (44,45).

Note: Do not use dUMP-containing primers with any archaeobacterial polymerase (*Pfu* DNA Polymerase, *Pwo* DNA Polymerase, etc.) or long PCR enzyme mixture.

2. Add the following to a 0.2 or 0.5-ml thin-wall PCR tube sitting on ice:

Component	Volume (µl)
sterilized, distilled water	28.5
10X PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl]	5.0
25 mM $MgCl_2$	3.0
10 mM dNTP mix	1.0
nested GSP2 (prepared as 10 µM solution)	1.0
AUAP (10 µM) or UAP (10 µM)	1.0
purified second-strand cDNA	10.0
final volume	49.5

Note: It is very important to have the reaction mixture ice-cold to avoid non-specific binding and extension of primers.

3. Add 0.5 µl of *Taq* DNA polymerase (5 units/µl) immediately before mixing.
4. Mix the contents of the tube and overlay with 50 to 100 µl of mineral oil (if necessary).

   Final composition of the reaction:

   20 mM Tris-HCl (pH 8.4)
   50 mM KCl
   1.5 mM $MgCl_2$
   200 nM GSP2
   200 nM UAP/AUAP
   200 µM each dATP, dCTP, dGTP, dTTP
   2.5 units *Taq* DNA polymerase

Note: The cycling protocol and annealing/extension temperatures may need to be optimized for each target sequence and/or thermal cycler. See section 2.2.7.

5. Transfer tubes directly from ice to the thermal cycler preequilibrated to the initial denaturation temperature (94°C).
6. Perform 30 to 35 cycles of PCR.

Seikagaku Corporation Exhibit 2004 - Page 140 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Additional Protocols

Note: The procedure requires careful pipetting of all solutions so that the concentration of divalent metal cation (Mg^{2+} and Ca^{2+}) is precise. Because the DNase I must be heated to 65°C to inactivate the enzyme, the concentration of free divalent metal ion must be low enough (< 1 mM) after the addition of the EDTA to prevent chemical hydrolysis of the RNA.

7. Analyze 5 to 20 µl of the amplified sample, using agarose gel electrophoresis, ethidium bromide staining, and the appropriate molecular size standards. If the positive control RNA was used, a prominent 711-bp band will be visible.

6.3 DNase I Digestion of RNA Preparation

If amplification products are detected from the PCR in the absence of SuperScript™ II RT, it may be necessary to eliminate residual genomic DNA from the RNA sample. After confirming the efficiency of the first strand synthesis reaction with the control RNA, use the following protocol to remove genomic DNA from the total RNA preparation. Amplification Grade DNase I has been extensively purified to remove trace ribonuclease activities.

1. Add the following to a 0.5-ml microcentrifuge tube on ice:

Component	Volume (µl)
total RNA	1-2 µg
10X reaction buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl, 20mM $MgCl_2$]	1 µl
Amplification grade DNase I (1-2 units/µl)	1 µl
DEPC-treated water	to 10 µl

2. Incubate at room temperature for 15 min.
3. Terminate the reaction by addition of 1 µl of 25 mM EDTA.
4. Incubate for 15 min at 65°C to heat inactivate the DNase I, then place on ice for 1 min. Collect the reaction by brief centrifugation. This mixture can be used directly for reverse transcription.

6.4 Lithium Chloride Purification of RNA Preparation

If you obtained poor first strand cDNA synthesis from your RNA preparation, the RNA may have coprecipitated with polysaccharides or small RNAs (tRNA and 5S RNA) during the RNA isolation procedure. You may be able to recover highly purified mRNA using the following protocol adapted from Sambrook *et al.* (19).

1. To the RNA sample in DEPC-treated water, add 0.1 volume 8 M LiCl (RNase-free) and vortex. Incubate on ice for 2 h.
2. Centrifuge at 14,000 x *g* for 30 min at 4°C.
3. Remove the supernate, being careful not to disturb the pellet (the pellet may be difficult to see). Dissolve the pellet in 200 µl of DEPC-treated water by drawing the pellet in and out of a sterilized pipet tip.
4. Repeat steps 1 through 3.
5. Add 0.1 volume of 3 M NaOAc (pH 5.2) and 2 volumes of absolute (100%) ethanol (–20°C). Place the tube at –20°C for 30 min. Centrifuge at 14,000 x *g* for 30 min at 4°C.
6. Remove the supernate carefully. Overlay the pellet with 100 µl of 70% ethanol (–20°C) and centrifuge at 14,000 x *g* for 10 min at 4°C. Remove the supernate, and air dry the RNA pellet at room temperature.
7. Dissolve the pellet in 10 to 100 µl of DEPC-treated water.
8. If your starting cell or tissue sample was small (10^3 to 10^5 cells), dissolve the pellet in 13 µl of DEPC-treated water and use this entire amount for first strand synthesis. If your starting sample was large (10^6 to 10^7 cells), spectrophotometrically determine the concentration of total RNA recovered in step 5 before proceeding to section 3.5

Seikagaku Corporation Exhibit 2004 - Page 141 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

6.5 Amplification from an Agarose Plug

This procedure is intended to provide added specificity to the amplification procedure prior to cloning or sequencing (1,18) 5' RACE products. The following protocol, adapted from Buck and Axel (63), should be performed after completion of nested amplifcation (section 3.9), but prior to performing the cloning procedures described below.

1. Analyze 5-20 µl of 5' RACE products by agarose gel electrophoresis according to standard protocols, using appropriate size standards (19). Either TAE or TBE electrophoresis buffer may be used for the procedure. The volume of the sample used for analysis will depend on the volume and thickness of the sample well. Ultraviolet (UV) visualization of ethidium bromide-stained products should be performed using either a long wave UV or 302-nm wavelength source to minimize DNA nicking if products will be extracted for reamplification.
2. Using a capillary pipette, remove a small (1-5 µl) agarose plug from the area of the gel containing the band of interest.
3. Transfer the agarose plug to a 0.5-ml microcentrifuge tube.
4. Amplify as described in section 3.9. Added specificity may be gained from the procedure by performing a nested PCR using a second nested GSP (composed of cDNA sequences 3' of the original GSP2) and the UAP. If sequence data for a nested GSP is unavailable, re-amplification of the agarose plug using GSP2 and the UAP should still provide sufficient enrichment of specific 5' RACE products for cloning.

Note: Do not carry over any mineral oil from the PCR.

6.6 Proteinase K Treatment of 5' RACE Product

If you obtained poor cloning efficiency following amplification, residual bound *Taq* DNA polymerase may have hindered restriction endonuclease digestion, resulting in poor vector/insert ligation. Use the following protocol to ensure complete removal of *Taq* DNA polymerase after amplification (42).

1. Extract the sample once with phenol:chloroform (1:1) and once with chloroform.
2. Add 0.1 volume of proteinase K solution [500 µg/ml proteinase K, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA].
3. Incubate for 30 min at 37°C, and then for 10 min at 65°C.
4. Extract once with phenol:chloroform (1:1) and once with chloroform.
5. Transfer the aqueous layer to a fresh 1.5-ml microcentrifuge tube.
6. Ethanol-precipitate by addition of 0.1 volume of 3M NaOAc (pH 5.2) and 2 volumes of absolute ethanol (–20°C). Place tube at –20°C for 30 min.
7. Centrifuge at 13,000 x *g* for 30 min. at 4°C. Remove supernatant, rinse pellet with 70% ethanol and air-dry the pellet at room temperature.
8. Resuspend the sample in 10 µl of TE buffer before advancing to cloning procedures.

6.7 Cloning Using T4 DNA Polymerase

Another alternative to conventional cloning methods uses the 3' to 5' exonuclease activity of T4 DNA polymerase as the basis for cloning as described by Stoker (51). In this procedure, PCR products from the primary PCR with the Abridged Anchor Primer, or nested amplification reaction primed with the AUAP, are treated with T4 DNA polymerase to generate a *Not* I 5' overhang. Similarly, the user may design a site into their GSP2 (see section 3.4.2.). Figure 9 depicts this procedure using an *Acc* I-compatible end-derived from the GSP2. Alternative restriction endonucleases that yield ends compatible with either *Not* I or *Acc* I are shown in table 2.

Seikagaku Corporation Exhibit 2004 - Page 142 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Additional Protocols

Table 2. Protruding Ends Compatible with *Acc* I or *Not* I.

Protruding End	Unambiguous			Ambiguous
5´-CG	*Aci* I	*Hin*P1 I	*Nsp*V	*Acc* I
		Hpa II	*Psp*1406 I	
		Mae II	*Taq* I	
	*Bsa*H I	*Msp* I		
	Cla I	*Nar* I		
5´-GGCC	*Eae* I	*Not* I	*Xma* III	

1. Following amplification with the Abridged Anchor Primer, or nested PCR with the AUAP, remove dNTPs from the reaction by agarose gel electrophoresis. Resuspend 200 to 500 ng of the reaction in 34 µl of sterilized, distilled water.
2. To the purified amplification product, add the following:

Component	Volume (µl)
10X Tris-acetate buffer [0.33 M Tris-acetate (pH 8.0), 0.66 M potassium acetate, 0.1 M magnesium acetate, 5 mM DTT, 1 mg/ml BSA]	5.0
dTTP (1 mM)	5.0
dATP (1 mM)	5.0
T4 DNA polymerase (1 unit/µl)	1.0

3. Incubate the reaction at 37°C for 15 min. Heat inactivate the enzyme by incubating at 65°C for 10 min. Ethanol precipitate the reaction (19) and resuspend in 5 µl of TE Buffer.
4. Prepare the vector for ligation by digestion with *Acc* I and *Not* I (19). Purify the vector from the excised fragment by agarose gel electrophoresis.
5. Combine ~100 ng of nondephosphorylated *Not* I-*Acc* I-cut vector with 500 ng of the T4 DNA polymerase-treated 3' RACE product and appropriate ligation reaction buffer to a final volume of 20 µl. Add 1 or 2 units of T4 DNA ligase, and incubate at 20°C to 25°C for 2 to 4 h.
6. Dilute the ligation reaction 1:10 in TE [10 mM Tris-HCl (pH 7.5), 1 mM EDTA]. Transform competent cells such as MAX Efficiency® or Library Efficiency® DH5α using 1 to 2 µl of the vector-ligated DNA.

Note: For additional information about ligation reactions using T4 DNA ligase, see reference 74.

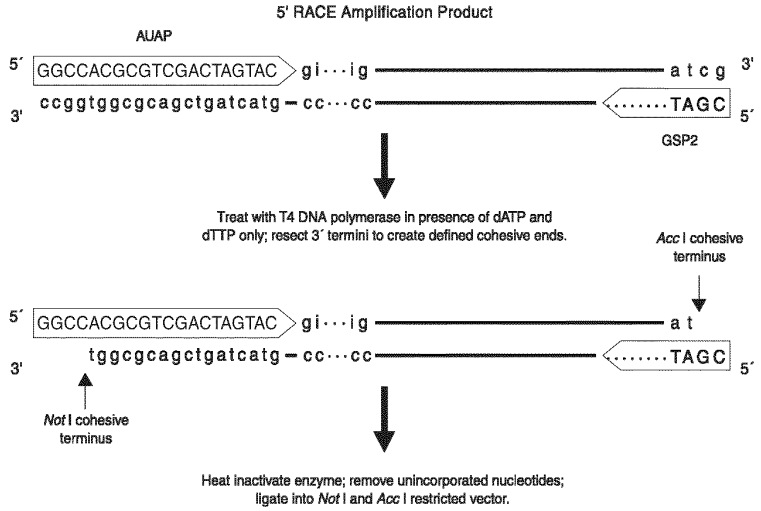


Figure 9: Cloning of 5' RACE Products using T4 DNA Polymerase.

Seikagaku Corporation Exhibit 2004 - Page 143 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

References

7

1. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA 85*, 8998.
2. Ohara, O., Dorit, R.L., and Gilbert, W. (1989) *Proc. Natl. Acad. Sci USA 86*, 5673.
3. Loh, Y., Elliott, J.F., Cwirla, S., Lanier, L.L., and Davis, M.M. (1989) *Science 243*, 217.
4. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985) *Science 230*, 1350.
5. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) *Science 239*, 487.
6. Frohman, M.A. (1990) PCR Protocols: A Guide to Methods and Applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds.) p. 28, Academic Press, San Diego.
7. Loh, E. (1991) Methods 2, 11.
8. Frohman, M.A., (1993) Rapid Amplification of Complementary DNA Ends for Generation of Full-Length Complementary DNAs: Thermal RACE. *Methods in Enzymology 218*:340-356.
9. Berchtold, M.W. (1989) *Nucleic Acids Res. 17*, 453.
10. Harvey, R.J., and Darlison, M.G. (1991) Nucleic Acids Res. 19, 4002.
11. Brow, M.A.D. (1990) PCR Protocols: A Guide to Methods and Applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds.) p. 189, Academic Press, San Diego.
12. Adams, S.M. and Blakesley, R. (1991) *Focus*® 13, 56.
13. Frohman, M.A. and Martin, G.R. (1989) *Techniques 1*, 165.
14. Buckler, A.J., Chang, D.D., Graw, S.L., Brook, D., Haber, D.A., Sharp, P.A., and Housman, D.E. (1991) *Proc. Natl. Acad. Sci USA 88*, 4005.
15. Nisson, P.E., Rashtchian, A., and Watkins, P.C. (1991) *PCR Meth. and Appl. 1*, 120.
16. Buchman, G.W. and Rashtchian, A. (1992) *Focus* 14, 2.
17. Buchman, G.W., Schuster, D.M., and Rashtchian, A. (1992) *Focus* 14, 41.
18. Craxton, M. (1991) *Methods 3*, 20.
19. Sambrook J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
20. Gerard, G.F., D'Alessio, J.M., and Kotewicz, M.L. (1989) *Focus* 11, 66.
21. Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem. 162*, 156.
22. Rychlik, W. and Rhodes, R.E. (1989) *Nucleic Acids Res. 17*, 8543.
23. Rychlik, W. (1995) *Biotechniques 18*,84.
24. Delidow, B. C., Lynch, J. P., Peluso, J. J. and White (1993) "Polymerase Chain Reaction" in PCR Protocols, Methods in Molecular Biology Vol. 15,1, B. A White, B. A. (Ed.) Humana Press Inc. Totowa, NJ.
25. Rychlik, W. (1993) Selection of Primers for Polymerase Chain Reaction" in PCR Protocols, Methods in Molecular Biology Vol. 15,31, B. A White, B. A. (Ed.) Humana Press Inc. Totowa, NJ.

Seikagaku Corporation Exhibit 2004 - Page 144 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

References

26. Innis, M. A. and Gelfand, D. (1990) "Optimization fo PCRs" in PCR Protocols, Innis, M.A. et al ,Eds., Academic Press, Inc., San Diego, p 3.
27. Martin, F.H., Castro, M.M., Aboul-ela, F., and Tinoco, I. Jr. (1985) *Nucleic Acids Res. 13*, 8927.
28. Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci USA 91*:2216-2220
29. Cheng, S., Fockler. C., Barnes, W.M., and Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA *Proc. Natl. Acad. Sci USA 91*:5695-5699.
30. D'Alessio, J.M., Gruber, C.E., Cain, C., and Noon, M.C. (1990) *Focus 12*, 47.
31. Gerard, G.F., Schmidt, B.J., Kotewicz, M.L., and Campbell, J.H. (1992) *Focus 14*, 91.
32. Ratliff, R.L. (1981) The Enzymes vol. XIV, (Boyer, P. ed), p 105, Academic Press, New York.
33. Pikaart, M.J., and Villeponteau, B. (1993) *Biotechniques 14*, 24.
34. Lobban, P.E. and Kaiser, A.D. (1973) *J. Mol. Biol. 78*, 453.
35. Nelson, T. and Brutlag, D., (1979) *Methods Enzymol. 68*, 41.
36. Schuster, D.M., Buchman, G.W., and Rashtchian, A. (1992) *Focus 14*, 46.
37. Vogelstein, B. and Gillespie, D. (1979) *Proc. Natl. Acad. Sci USA 76*, 615.
38. Saiki, R.K. (1989) PCR Technology, Principles and Applications for DNA Amplification (Erlich, H.A., ed) p 17, Stockton Press, New York.
39. Eckert, K.A. and Kunkel, T.A. (1991) *PCR Meth. and Appl. 1*, 17.
40 Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C., and Sninsky, J.J. (1990) *Nucleic Acids Res. 18*, 999.
41. Gustafson, C.E., Alm, R.A., and Trust, T.J., (1993) Effect of heat denaturation of target DNA on the PCR amplification. *Gene 123*, 241.
42. Innis, M.A. and Gelfand, D.H. (1990) PCR Protocols: A Guide to Methods and Applications (Innis, M.A., Gelfand, D. H., Sninsky, J.J., and White, T.J., eds) p. 3, Academic Press, San Diego.
43. Saiki, R.K. (1990) PCR Protocols: A Guide to Methods and Applications (Innis, M.A., Gelfand, D. H., Sninsky, J.J., and White, T.J., eds) p. 13, Academic Press, San Diego.
44. D'Aquila, R.T., Bechtel, L.J., Videler, J.A., Eron, J.J., Gorczyca, P., and Kaplan, J.C. (1991) *Nucleic Acids Res. 19*, 3749.
45. Mullis, K.B. (1991) *PCR Meth. and Appl. 1*, 1.
46. Bhat, G.J., Lodes, M. J., Myler, J., and Stuart, K.D. (1991) *Nucleic Acids Res. 19*, 398.
47. Shuldiner, A.R., Scott, L.A., and Roth, J. (1990) *Nucleic Acids Res. 18*, 1920.
48. Rashtchian, A. (1995) Novel methods for cloning and engineering genes using the polymerase chain reaction. *Current Opinion in Biotechnology 6,* 30.
49. Witcomb, J.M., Rashtchian, A., Hughes, S.H., 1993, A new method for the generation of nested deletions. *Nucleic Acids Research 21*, 4143.
50. Carney, J.P., McKnight, C., Van Epps S, Kelly, M.R. (1995) Random rapid amplification of cDNA ends (RRACE) allows for cloning of multiple novel human cDNA fragments containing (CAG)n repeats. *Gene 155*, 289.
51. Stoker, A. W. (1990) *Nucleic Acids Res. 18*, 4290.

Seikagaku Corporation Exhibit 2004 - Page 145 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

52. Gerard, G. (1995) *Focus 16*, 102.

53. Chomczynski, P. (1993) *Biotechniques 15*, 532.

54. Simms, D., Cizdziel, P. E. and Chomczynski, P. (1993) *Focus 11*, 99.

55. Saiki, R.K. (1989) PCR Technology: Principles and Applications for DNA Amplification (Erlich, H.A., ed.) p 17 Stockton Press, New York

56. Rychlik, W., Spencer, W.J., and Rhodes, R.E. (1990) *Nucleic Acids Res. 18*, 6409.

57. Lowe, T., Sharefkin, J., Yang, S.Q., and Diffenbach, C.W. (1990) Nucleic Acids Res. 18, 1757.

58. Hiller, L. and Green, P. (1991) *PCR Meth. and Appl. 1*, 124.

59. Scharf, S.J. (1990) PCR Protocols: A Guide to Methods and Applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds) p. 84, Academic Press, San Diego.

60. Knoth, K., Roberds, S., Poteet, C., and Tamkun, M. (1988) *Nucleic Acids Res. 16*, 10932.

61. Patil, R.V. and Dekker, E.E. (1990) *Nucleic Acids Res. 18*, 3080.

62. Fordham-Skelton, A.P., Yarwood, A., and Croy, R.R.D. (1990) *Mol. Gen. Genet. 221*, 134.

63. Buck. L. and Axel, R. (1991) *Cell 65*, 175.

64. Crowe, J.S., Cooper, H.J., Smith, M.A., Sims, M.J., Parker, D., and Gewert, D. (1991) *Nucleic Acids Res. 19*, 184.

65. Breslauer, K.L., Frank, R.H., and Marky, L.A. (1986) *Proc. Natl. Acad. Sci USA 83*, 3746.

66. Mueller, H.W. and Duclos, J.M. (1993) *Amplifications 11*, 10.

67. Itakura, K., Rossi, J.J., and Wallace, R.B. (1984) *Ann. Rev. Biochem. 53*, 323.

68. Suggs, S.V., Hirose, T., Miyake, T., Kawashima, E.H., Johnson, M.J., Itakura, K., and Wallace, R.B. (1981) In: Developmental Biology Using Purified Genes, eds. Brown, D.D. and Fow, C.F. Academic Press, New York.

69. Baldino, F., Chesselet, M-F., and Lewis, M.E. (1989) *Meth. Enz. 168*, 761.

70. Bolton, E.T. and McCarthy, B.J. (1962) *Proc. Natl. Acad. Sci USA 48*, 1390.

71. Life Technologies, Inc. (1990) *Focus on Applications: T4 DNA Ligase (Technical Bulletin 5224-1).*

Seikagaku Corporation Exhibit 2004 - Page 146 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

8 Related Products

Product	Size	Cat. No.
Primers:		
Custom Primers		www.invitrogen.com
Abridged Universal Amplification Primer	2.5 µg	10541-019
Universal Amplification Primer	5 µg	18382-010
5´ RACE Abridged Anchor Primer	5.7 µg	10630-010
5´ RACE Anchor Primer	7.5 µg	18388-017
3´ RACE Adapter Primer	2.3 µg	10542-017
Restriction Endonucleases:		
Mlu I	1,000 units	15432-016
Mlu I High Concentration	5,000 units	15432-024
Sal I	1,000 units 5,000 units	15217-011 15217-029
Sal I High Concentration	5,000 units	15217-037
Spe I	200 units 800 units (4 × 200 units)	15443-013 15543-047
Products for Transformation and Cloning:		
MAX Efficiency® DH5α™ Competent Cells	1 ml (5 × 0.2 ml)	18258-012
MAX Efficiency® DH10B™ Competent Cells	1 ml (5 × 0.2 ml)	18297-010
Library Efficiency® DH5α™ Competent Cells	1 ml (5 × 0.2 ml)	18263-012
MAX Efficiency® Stbl2™ Competent Cells	1 ml (5 × 0.2 ml)	10268-019
S.O.C. Medium	10 × 10 ml	15544-018
Enzymes for Amplification:		
Taq DNA Polymerase, Native	100 units 500 units 1,500 units (3 × 500 units)	18038-018 18038-042 18038-067
Taq DNA Polymerase, Recombinant	100 units 500 units 1,500 units (3 × 500 units)	10342-053 10342-020 10342-046
Platinum® *Taq* DNA Polymerase	250 units	10966-026
eLONGase® Amplification System	100 reactions	10481-018
eLONGase® Enzyme Mix	100 reactions 500 reactions	10480-010 10480-028

Seikagaku Corporation Exhibit 2004 - Page 147 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Product	Size	Cat. No.
Other Modifying Enzymes:		
Proteinase K	100 mg 1 g	25530-015 25530-031
Proteinase K Solution	5 ml (100 mg)	25530-049
T4 DNA Ligase	100 units 500 units 2,000 units (4 × 500 units)	15224-017 15224-025 15224-090
T4 DNA Ligase High Concentration	250 units	15224-041
DNase I, Amplification Grade	100 units	18068-015
SuperScript™ II Reverse Transcriptase	2,000 units 10,000 units 40,000 units (4 × 10,000 units)	18064-022 18064-014 18064-071
RNaseOUT™, Recombinant RNase Inhibitor	5,000 units	10777-019
Other Related Products:		
TRIzol® Reagent	100 ml 200 ml	15596-026 15596-018
3′ RACE System for Rapid Amplification of cDNA Ends	20 reactions	18373-019
100 bp DNA Ladder	50 µg 250 µg	15628-019 15628-050
1 Kb DNA Ladder	250 µg 1 mg	15615-016 15615-024
DEPC-Treated Water	4 × 1.25 ml	10813-012

Seikagaku Corporation Exhibit 2004 - Page 148 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Notes

Seikagaku Corporation Exhibit 2004 - Page 149 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023



United States Headquarters:
Invitrogen Corporation
1600 Faraday Avenue
Carlsbad, California 92008
Tel: 1 760 603 7200
Tel (Toll Free): 1 800 955 6288
Fax: 1 760 603 7229
Email: tech_service@invitrogen.com

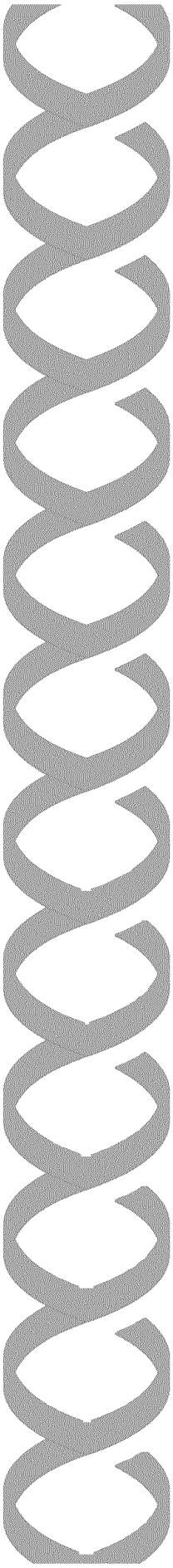
European Headquarters:
Invitrogen Ltd
3 Fountain Drive
Inchinnan Business Park
Paisley PA4 9RF, UK
Tel (Free Phone Orders): 0800 269 210
Tel (General Enquiries): 0800 5345 5345
Fax: +44 (0) 141 814 6287
Email: eurotech@invitrogen.com

International Offices:
Argentina 5411 4556 0844
Australia 1 800 331 627
Austria 0800 20 1087
Belgium 0800 14894
Brazil 0800 11 0575
Canada 800 263 6236
China 10 6849 2578
Denmark 80 30 17 40

France 0800 23 20 79
Germany 0800 083 0902
Hong Kong 2407 8450
India 11 577 3282
Italy 02 98 22 201
Japan 03 3663 7974
The Netherlands 0800 099 3310
New Zealand 0800 600 200
Norway 00800 5456 5456

Spain & Portugal 900 181 461
Sweden 020 26 34 52
Switzerland 0800 848 800
Taiwan 2 2651 6156
UK 0800 838 380
For other countries see our website

www.invitrogen.com



Invitrogen™
life technologies

Instruction Manual

3´ RACE System for Rapid Amplification of cDNA Ends

Catalog no. 18373-019

Version E
6 December 2004
10448

Table of Contents

1. Notices to Customer 1
- 1.1 Notice to Customer 1
- 1.2 Precautions 1
- 1.3 Limited Label Licenses 1

2. Overview 2
- 2.1 Principles of RACE 2
- 2.2 Summary of the 3′ RACE System 2
- 2.3 Isolation of Total RNA 3
- 2.4 First Strand cDNA Synthesis from Total RNA 4
- 2.5 Amplification of a Target cDNA 5
- 2.6 Design of the Gene-Specific Primer 5
- 2.7 Nested Amplification 5
- 2.8 Cloning of Amplification Products 6

3. Methods 7
- 3.1 Components 7
- 3.2 Advance Preparations 7
- 3.3 Protocol 1. First Strand cDNA Synthesis 8
- 3.4 Protocol 2. Amplification of the Target cDNA 8

4. Troubleshooting Guide 10
- 4.1 General Suggestions 10
- 4.2 Testing the 3′ RACE System 11
  - 4.2.1 First Strand cDNA Synthesis 11
  - 4.2.2 Amplification of the Control-Synthesized cDNA 12
- 4.3 Minimizing RNase Contamination 13

5. Additional Protocols 15
- 5.1 First Strand cDNA Synthesis of Transcripts with High GC Content 15
- 5.2 TRIzol® Reagent Isolation of Total RNA 16
- 5.3 DNase I Digestion of RNA Preparation 17
- 5.4 Lithium Chloride Purification of RNA Preparation 17
- 5.5 Nested Amplification from an Agarose Plug 18
- 5.6 Cloning Using T4 DNA Polymerase 19
- 5.7 Proteinase K Treatment of 3′ RACE Products 19

6. References 20

7. Related Products 21

Seikagaku Corporation Exhibit 2004 - Page 153 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Table of Contents

Figures

1. Summary of the 3´ RACE System Procedure 3
2. AP, UAP, and AUAP Primer Sequences 4
3. Sequence of the Control GSP 12
4. Agarose Gel Analysis of 3´ RACE Products Using Control RNA 13

Table

1. Protruding Ends Compatible with *Acc* I or *Not* I 6

TRIzol® is a registered trademark of Molecular Research Center, Inc.

Teflon® is a registered trademark of E.I. du Pont de Nemours & Co.

Seikagaku Corporation Exhibit 2004 - Page 154 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Notices to Customer

1

1.1 Important Information

This product is authorized for laboratory research use only. The product has not been qualified or found safe and effective for any human or animal diagnostic or therapeutic application. Uses for other than the labeled intended use may be a violation of applicable law.

1.2 Precautions

Warning: This product contains hazardous reagents. It is the end-user's responsibility to consult the applicable MSDS(s) before using this product. Disposal of waste organics, acids, bases, and radioactive materials must comply with all appropriate federal, state, and local regulations. If you have any questions concerning the hazards associated with this product, please call the Invitrogen Corp. Environmental Health and Safety Chemical Emergency hotline at (301) 431-8585.

1.3 Limited Label Licenses

Limited Use Label License No. 4: Products for PCR that include no rights to perform PCR

This product is optimized for use in the Polymerase Chain Reaction (PCR) covered by patents owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche, Ltd. ("Roche"). No license under these patents to use the PCR process is conveyed expressly or by implication to the purchaser by the purchase of this product. A license to use the PCR process for certain research and development activities accompanies the purchase of certain reagents from licensed suppliers such as Invitrogen, when used in conjunction with an Authorized Thermal Cycler, or is available from Applied Biosystems. Further information on purchasing licenses to practice the PCR process may be obtained by contacting the Director of Licensing at Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404 or at Roche Molecular Systems, Inc., 1145 Atlantic Avenue, Alameda, California 94501.

Limited Label License No. 5: SuperScript™ Reverse Transcriptase

The purchase of this product conveys to the buyer the non-transferable right to use the purchased amount of the product and components of the product in research conducted by the buyer (whether the buyer is an academic or for-profit entity). The buyer cannot sell or otherwise transfer (a) this product (b) its components or (c) materials made using this product or its components to a third party or otherwise use this product or its components or materials made using this product or its components for Commercial Purposes. The buyer may transfer information or materials made through the use of this product to a scientific collaborator, provided that such transfer is not for any Commercial Purpose, and that such collaborator agrees in writing (a) not to transfer such materials to any third party, and (b) to use such transferred materials and/or information solely for research and not for Commercial Purposes. Commercial Purposes means any activity by a party for consideration and may include, but is not limited to: (1) use of the product or its components in manufacturing; (2) use of the product or its components to provide a service, information, or data; (3) use of the product or its components for therapeutic, diagnostic or prophylactic purposes; or (4) resale of the product or its components, whether or not such product or its components are resold for use in research. Invitrogen Corporation will not assert a claim against the buyer of infringement of patents owned by Invitrogen and claiming this product based upon the manufacture, use or sale of a therapeutic, clinical diagnostic, vaccine or prophylactic product developed in research by the buyer in which this product or its components was employed, provided that neither this product nor any of its components was used in the manufacture of such product. If the purchaser is not willing to accept the limitations of this limited use statement, Invitrogen is willing to accept return of the product with a full refund. For information on purchasing a license to this product for purposes other than research, contact Licensing Department, Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, California 92008. Phone (760) 603-7200. Fax (760) 602-6500.

Limited Use Label License No. 43: Universal Amplification Primer

This product is the subject of U.S. Patent No. 5,334,515 owned by Invitrogen Corporation. The purchase of this product conveys to the buyer the non-transferable right to use the purchased amount of the product and components of the product in research conducted by the buyer (whether the buyer is an academic or for-profit entity). The buyer cannot sell or otherwise transfer (a) this product (b) its components or (c) materials made using this product or its components to a third party or otherwise use this product or its components or materials made using this product or its components for Commercial Purposes. The buyer may transfer information or materials made through the use of this product to a scientific collaborator, provided that such transfer is not for any Commercial Purpose, and that such collaborator agrees in writing (a) to not transfer such materials to any third party, and (b) to use such transferred materials and/or information solely for research and not for Commercial Purposes. Commercial Purposes means any activity by a party for consideration and may include, but is not limited to: (1) use of the product or its components in manufacturing; (2) use of the product or its components to provide a service, information, or data; (3) use of the product or its components for therapeutic, diagnostic or prophylactic purposes; or (4) resale of the product or its components, whether or not such product or its components are resold for use in research. Invitrogen Corporation will not assert a claim against the buyer of infringement of the above patent based upon the manufacture, use or sale of a therapeutic, clinical diagnostic, vaccine or prophylactic product developed in research by the buyer in which this product or its components was employed, provided that neither this product nor any of its components was used in the manufacture of such product. If the purchaser is not willing to accept the limitations of this limited use statement, Invitrogen is willing to accept return of the product with a full refund. For information on purchasing a license to this product for purposes other than research, contact Licensing Department, Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, California 92008. Phone (760) 603-7200. Fax (760) 602-6500.

Seikagaku Corporation Exhibit 2004 - Page 155 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

2 Overview

2.1 Principles of RACE

Rapid Amplification of cDNA Ends (RACE) is a procedure for amplification of nucleic acid sequences from a messenger RNA template between a defined internal site and either the 3′ or the 5′ end of the mRNA (1). This methodology of amplification with single-sided specificity has been described as "one-sided" PCR (2) or "anchored" PCR (3). PCR requires two sequence-specific primers that flank the sequence to be amplified (4,5). However, to amplify and characterize regions of unknown sequences, this requirement imposes a limitation (3).

3′ RACE takes advantage of the natural poly(A) tail found in mRNA as a generic priming site for PCR. In this procedure, mRNAs are converted into cDNA using reverse transcriptase (RT) and an oligo-dT adapter primer. Specific cDNA is then amplified by PCR using a gene-specific primer (GSP) that anneals to a region of known exon sequences and an adapter primer that targets the poly(A) tail region. This permits the capture of unknown 3′-mRNA sequences that lie between the exon and the poly(A) tail.

5′ RACE uses an antisense gene specific primer for the synthesis of specific cDNA by reverse transcriptase. Prior to PCR, a TdT-tailing step attatches an adapter sequence to the unknown 5′ sequences of the cDNA. Specific cDNA is then amplified by PCR using a GSP that anneals in a region of known exon sequences and an adapter primer that targets the 5′ terminus.

RACE has been used for amplification and cloning of rare mRNAs (6) and may be applied to existing cDNA libraries (7). Additionally, RACE products can be directly sequenced without any intermediate cloning steps (8,9), or the products may be used to prepare probes (10). Products generated by the 3′ and 5′ RACE procedures may be combined to generate full-length cDNAs (10,11). Lastly, the RACE procedures may be utilized in conjunction with exon-trapping methods (12) to enable amplification and subsequent characterization of unknown coding sequences.

2.2 Summary of the 3′ RACE System

The 3′ RACE procedure is summarized in figure 1. First strand cDNA synthesis is initiated at the poly(A) tail of mRNA using the adapter primer (AP). After first strand cDNA synthesis, the original mRNA template is destroyed with RNase H, which is specific for RNA:DNA heteroduplex molecules. Amplification is performed, without intermediate organic extractions or ethanol precipitations, using two primers: one is a user-designed GSP that anneals to a site located within the cDNA molecule; the other is a universal amplification primer that targets the mRNA of the cDNA complementary to the 3′ end of the mRNA. Two universal amplification primers are provided with the system. The universal amplification primer (UAP) is designed for the rapid and efficient cloning of RACE products using the uracil DNA glycosylase (UDG) cloning method (13–16). The abridged universal amplification primer (AUAP) is homologous to the adapter sequence used to prime first strand cDNA synthesis.

Seikagaku Corporation Exhibit 2004 - Page 156 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

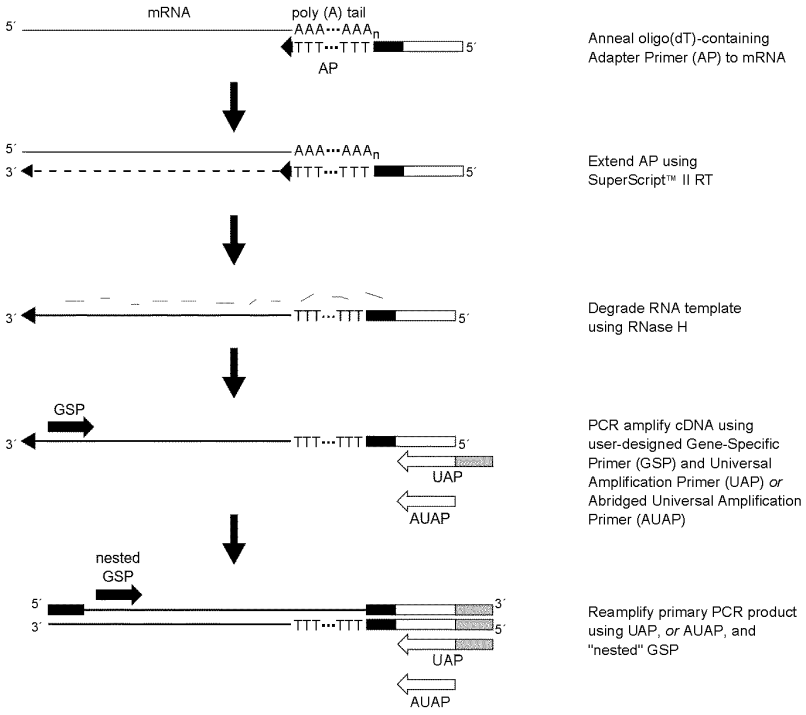


Figure 1. Summary of the 3´ RACE System procedure.

Since the 3´ RACE System utilizes the poly(A) tail region as an initial priming site, multiple amplification products may be synthesized, depending on the degree of specificity conferred by the GSP. To generate a specific amplification product, the user may find it advantageous to design a second "nested" GSP, as recommended by Frohman *et al.* (10) and reamplify the RACE products; this procedure is discussed in greater detail at the end of this chapter.

2.3 Isolation of Total RNA

One of the most important factors preceding the synthesis of substantially full-length cDNA is the isolation of intact RNA. The quality of the RNA dictates the maximum amount of sequence information that can be converted into cDNA. Thus, it is important to optimize the isolation of RNA from a given biological source and to prevent adventitious introduction of RNases (17) and inhibitors of reverse transcriptase such as guanidinium salts, SDS and EDTA (18). RNA can be isolated using a variety of methods. The recommended method for 3´ RACE is the guanidine isothiocyanate/acid-phenol method originally described by Chomzynski and Sacchi (19). The TRIzol® Reagent method is an improvement of the original single-step method of Chomczynski and Sacchi (20) and can be used for the preparation of RNA from as little as 10^3 cells or milligram quantities of tissue (21). Total RNA isolated with TRIzol® Reagent is undegraded and essentially free of protein and DNA contamination. The TRIzol® RNA isolation protocol is described in Section 6.2. For the isolation of RNA from small quantities of sample (<10^6 cells or <10 mg tissue) without using phenol, the GlassMAX RNA Microisolation Spin Cartridge System is recommended (22).

Seikagaku Corporation Exhibit 2004 - Page 157 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Total RNA isolated by these methods may contain small amounts of genomic DNA that may subsequently be amplified along with the target cDNA. The presence of this DNA is not likely to cause problems because it lacks the poly(A) region present in the mRNA analyte. As a precaution, however, we recommend performing a control experiment without reverse transcriptase to determine whether a given fragment is of genomic DNA or of cDNA origin. Products generated in the absence of RT are of genomic origin. If your application requires removal of all genomic DNA from your RNA preparation, refer Section 5.3, *DNase I Digestion of RNA Preparation.*

2.4 First Strand cDNA Synthesis from Total RNA

The first strand cDNA synthesis reaction is catalyzed by SuperScript™ II RT. This enzyme is a mutant of M-MLV RT that has been engineered to reduce RNase H activity, resulting in greater yields and more full-length synthesis (23,24,25). The enzyme exhibits increased thermal stability and may be used at temperatures up to 50°C. In addition, SuperScript™ II RT is not inhibited significantly by ribosomal and transfer RNA and may be used to synthesize first strand cDNA from a total RNA preparation. The RNA template is removed from the cDNA:RNA hybrid molecule by digestion with RNase H after cDNA synthesis to increase the sensitivity of PCR (26).

The AP (see figure 2), which primes first strand cDNA synthesis, has been engineered to contain three restriction endonuclease sites and a *Not* I half-site. Inclusion of these sequences in the primer may facilitate post-amplification cloning using either a restriction endonuclease-based (27) or a T4 DNA polymerase-based (28) method. Because the AP initiates cDNA synthesis at the poly(A) region of the mRNA, it effectively selects for polyadenylated mRNAs; thus, oligo(dT)-selection for poly(A)$^+$ RNA is typically not necessary although incorporating this step may facilitate the detection of rare mRNA transcripts.

2.5 Amplification of a Target cDNA

Amplification of a target cDNA requires priming with two oligonucleotides and *Taq* DNA polymerase. The sense amplification primer is the user-provided GSP,

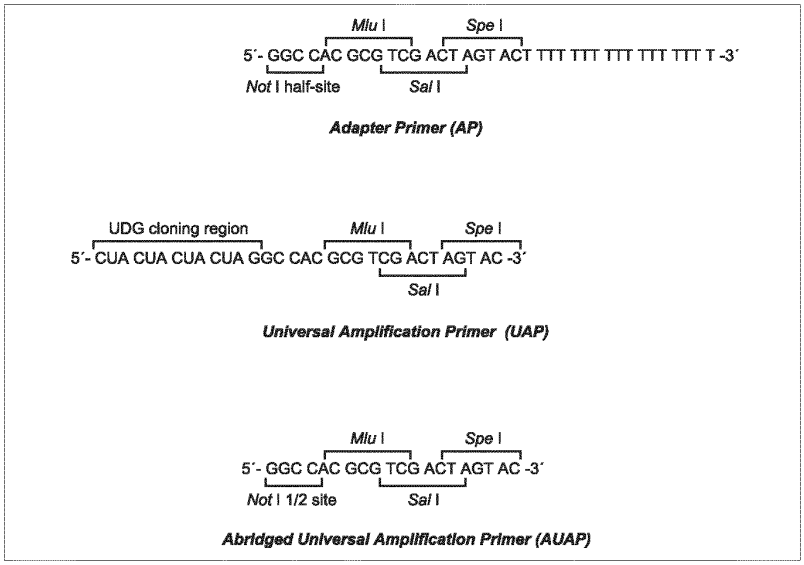


Figure 2. AP, UAP, and AUAP primer sequences.

Seikagaku Corporation Exhibit 2004 - Page 158 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

which is specific for the particular gene or sequence of interest and may be designed to include sequence elements that facilitate subsequent cloning steps. The antisense amplification primer is one of the two universal amplification primers (see figure 2) provided with the system. The AUAP contains a restriction endonuclease site sequence (adapter region) homologous to the adapter region of the AP. The UAP is composed of the same adapter region plus a dUMP-containing sequence at the 5′ end of the primer required for UDG-mediated cloning. The UAP should not be used to prime DNA synthesis with any archaeobacterial polymerase (*e.g., Pyrococcus furiosus, Pyrococcus woesei,* etc.) or any long PCR enzyme mixture (*e.g.,* Elongase® Enzyme Mix) that contains one of these enzymes due to the inhibition of polymerase activity by dUMP-containing DNA. Both the AUAP and the UAP will function in PCR at annealing temperatures up to 68°C.

2.6 Design of the Gene-Specific Primer

Efficient and specific PCR amplification is highly dependent on primer design. This is especially true for RACE applications since the PCR is carried out with only a single GSP. In general, effective primers form stable duplexes with their target sequences, are highly specific for their target sequences, and are free of secondary structure such as hairpin loops and dimers (29–31). Additionally, the complementarity of primer 3′-termini must be minimized since primer-dimer artifacts may significantly reduce PCR efficiency. Therefore, dimer formation with the AUAP or UAP primer, as well as itself, should be reduced. Computer algorithms that have been developed (32–35) and are commercially available often facilitate this analysis. Discussion of primer design for RACE applications may be found in Frohman (11) and Loh (6). It should be noted that in cases where only limited peptide sequence information is available, a degenerate GSP may be prepared.

The AUAP and UAP included with the system have been engineered to function at PCR annealing temperatures up to 68°C and to facilitate the cloning step. The user-defined GSPs need to be compatible with the cloning method. Add the following to the 5' end of the GSP:

for UDG cloning:	5′–CAU CAU CAU CAU–3′	(use with UAP)
for T4 DNA polymerase cloning:	5′–CGA–3′	(use with AUAP)

2.7 Nested Amplification

The AP is designed to synthesize first strand cDNA from all polyadenylated mRNAs. The sequence specificity in the amplification reaction is therefore derived solely from the GSP. Often, a second "nested" GSP may be utilized in conjunction with the AUAP or UAP in a second amplification reaction to give the 3′ RACE procedure the specificity of a second primer (9). The nested GSP can anneal immediately adjacent to the first GSP or at sequences within the cDNA further downstream. The nested amplification reaction may be conveniently conducted using a plug of agarose from the gel analysis of the initial 3′ RACE reaction (see Section 5.5, *Nested Amplification from an Agarose Plug*). Ultimately, the 3′ RACE procedure should produce a single, prominent band on an agarose gel. When performing 3′ RACE with a nested primer, sequences specific for subsequent cloning manipulations (see Section 2.6, *Design of the Gene-Specific Primer*) must be designed into the nested GSP.

Seikagaku Corporation Exhibit 2004 - Page 159 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

2.8 Cloning of Amplification Products

Conventional cloning methods that typically involve end-repair and blunt-end cloning can be problematic for amplified products (36–38). An alternative is a rapid and efficient method involving the use of UDG (13–16). This method requires that the user design a GSP containing containing a 5'-$(CAU)_4$ sequence. Incorporation of dUMP into the GSP may be accomplished on most automated synthesizers or with Invitrogen Custom Primers (see Section 2.6, *Design of the Gene-Specific Primer*). The product of the 3′ RACE reaction primed with the UAP and the dUMP-containing GSP is treated with UDG, which converts dUMP residues to abasic sites (39,40), to generate 3′ overhangs. The directional nature of the UDG cloning process can be exploited to lend an added level of specificity to the RACE procedure. Only amplification product that results from priming by both the UAP and the appropriately designed GSP are efficient substrates for UDG cloning.

Another alternative to conventional cloning methods uses the 3′ to 5′ exonuclease activity of T4 DNA polymerase as the basis for cloning as described by Stoker (28). In this procedure, the AUAP is used in the amplification reaction, and the 3′ RACE products are treated with T4 DNA polymerase to generate a *Not* I 5′ overhang. Similarly, the user may design a site into the GSP (see Section 2.6, *Design of the Gene-Specific Primer*).

Another approach to cloning is to digest the 3′ RACE product using one of the restriction endonuclease sites designed into the AUAP (figure 2; ref. 1). The user may also design unique restriction sites into the GSP, exploit a site present in the cDNA sequence or end-repair the 3′ RACE product prior to restriction endonuclease digestion (37).

Seikagaku Corporation Exhibit 2004 - Page 160 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Methods

3

3.1 Components

Components are provided in sufficient quantities to perform 20 separate reactions, each converting 1-5 µg of total RNA into first strand cDNA. A control RNA and amplification primers are included in the system to verify performance of the first strand cDNA synthesis reaction and subsequent amplification (see Chapter 5 for specific protocols). **Note:** The 3´ RACE System does not include *Taq* DNA polymerase or the reagents required for cloning. Store the 3´ RACE System at –20°C.

Component	Amount
10X PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl]	500 µl
25 mM $MgCl_2$	500 µl
10 mM dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)	100 µl
0.1 M DTT	100 µl
SuperScript™ II Reverse Transcriptase (RT, 200 units/µl)	20 µl
adapter primer (AP, 10 µM)	20 µl
universal amplification primer (UAP, 10 µM)	20 µl
abridged universal amplification primer (AUAP, 10 µM)	20 µl
E. coli RNase H (2 units/µl)	20 µl
DEPC-treated water	1.2 ml
control RNA (50 ng/µl)	10 µl
control gene-specific primer (GSP, 10 µM)	20 µl

3.2 Advance Preparations

Please review the protocols before using this system. You will need the following items *not included* in the system:

- sterilized 0.5-ml microcentrifuge tubes;
- automatic pipets capable of dispensing 1 to 20 µl and 20 to 200 µl;
- sterilized, RNase-free disposable tips for automatic pipets;
- disposable latex gloves;
- sterilized, distilled water;
- GSP (user-defined, appropriately engineered);
- microcentrifuge capable of generating a relative centrifugal force of 14,000 × *g;*
- 37°C, 42°C, and 70°C water baths or heat blocks;
- *Taq* DNA polymerase;
- chloroform;
- mineral oil; and
- thin-walled PCR tubes.

Seikagaku Corporation Exhibit 2004 - Page 161 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

3.3 Protocol 1. First Strand cDNA Synthesis

This procedure converts 1 to 5 µg of total RNA into first strand cDNA. Poly(A)$^+$ RNA may be used in this protocol, but is typically not necessary. A control RNA is included in the 3´ RACE System as an aid in verifying that the system performs in your hands. If you decide to use the control RNA as a template for first strand synthesis, simply substitute 2 µl of control RNA (100 ng) in the first strand reaction for your total RNA. Please see Chapter 5 for additional information.

Note: For the control RNA, use 2 µl (100 ng). See chapter 4 for additional information.

Note: If you have >5 µg of total RNA, increase reaction volumes and amount of SuperScript™ II RT proportionately. If you have <1 µg of total RNA, no changes to the protocol are necessary. 50 to 500 ng of poly(A)$^+$ RNA may be substituted for total RNA in this protocol.

1. Mix and quickly centrifuge each component before use.
2. Combine 1-5 µg of total RNA or 50 ng of poly(A)$^+$ RNA and DEPC-treated water to a final volume of 11 µl in a 0.5-ml microcentrifuge tube.
3. Add 1 µl of the 10 µM AP solution, mix gently, and collect reaction by brief centrifugation.
4. Heat the mixture to 70°C for 10 min and chill on ice for at least 1 min. Collect the contents of the tube by brief centrifugation and add the following:

Component	Volume (µl)
10X PCR buffer	2
25 mM $MgCl_2$	2
10 mM dNTP mix	1
0.1 M DTT	2

Final composition of the reaction:

20 mM Tris-HCl (pH 8.4 at 22°C)
50 mM KCl
2.5 mM $MgCl_2$
10 mM DTT
500 nM AP
500 µM each dATP, dCTP, dGTP, dTTP
1-5 µg (≤50 ng/µl) of RNA

5. Mix gently and collect the reaction by brief centrifugation. Equilibrate the mixture to 42°C for 2 to 5 min.
6. Add 1 µl of SuperScript™ II RT. Incubate the tube in a 42°C water bath or heat block for 50 min.
7. Terminate the reaction by incubating at 70°C for 15 min.
8. Chill on ice. Collect the reaction by brief centrifugation. Add 1 µl of RNase H to the tube, mix, and incubate for 20 min at 37°C before proceeding to Protocol 2.
9. The reaction mixture may be stored at –20°C.

Note: You may stop at the end of step 9.

3.4 Protocol 2. Amplification of the Target cDNA

Optimal conditions for amplification are dependent on the nature of each particular primer and target sequence used. Alteration of the magnesium ion, dNTP, or primer concentration, as well as the thermocycling protocol, may be required. The optimal free magnesium concentration for efficient amplification is reported to be between 0.7 to 0.8 mM (29). Since magnesium binds deoxyribonucleoside triphosphates, this factor is affected by both primer and dNTP concentration. In general, lower concentrations of dNTP (50 to 200 µM), $MgCl_2$ (1 to 1.5 mM), and primer (0.1 to 0.2 µM) promote higher fidelity and specificity (41). Higher nucleotide concentration, however, can be used to improve product yield as well as to promote 3´-terminal T-mismatches (42). For a detailed discussion of parameters affecting PCR, please refer to Innis and Gelfand (43) or Saiki (29,30).

Seikagaku Corporation Exhibit 2004 - Page 162 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

The addition of either *Taq* DNA polymerase, dNTPs, or $MgCl_2$ after reactions have been equilibrated at 75°C to 80°C has been reported to improve the specificity of the reaction (44,45). This "hot start" practice reduces nonspecific binding and the extension of primers during the initial denaturation process. A practical alternative to this classic "hot start" method is to set up reactions on ice then place complete PCR mixtures in a thermal cycler equilibrated to 80°C to 90°C.

1. To a fresh 0.5-ml microcentrifuge tube, add the following:

Component	Volume (µl)
10X PCR buffer	5
25 mM $MgCl_2$	3
autoclaved, distilled water	36.5
10 mM dNTP mix	1
GSP (prepared as 10 µM solution)	1
AUAP (10 µM) or UAP (10 µM)	1
Taq DNA polymerase (2 to 5 units/µl)	0.5

Note: Control cDNA should be diluted prior to amplification, see Section 4.2.2.

Final composition of the reaction:

20 mM Tris-HCl (pH 8.4)
50 mM KCl
1.5 mM $MgCl_2$
200 nM GSP
200 nM AUAP or UAP
200 µM each dATP, dCTP, dGTP, dTTP
0.04 to 0.1 unit/µl *Taq* DNA polymerase

Note: Mineral oil is necessary only when using thermal cyclers that require this barrier.

2. Add 2 µl from the cDNA synthesis reaction to the tube. Mix gently and layer 75 µl of mineral oil over the reaction. Collect the reaction briefly by centrifugation.
3. Incubate the reaction at 94°C for 3 min. (Extended preamplification denaturation times may impair the efficiency of long (>2 kb) target sequences (46).
4. Perform 20 to 35 cycles of PCR, using the protocol accompanying *Taq* DNA polymerase.

Note: Step 5 is unnecessary if no mineral oil was used in step 2.

5. Following amplification, extract the sample with 50 µl of chloroform. Transfer the aqueous layer to a fresh tube.

Note: If sequences are available for use as internal probes, it is strongly recommended that Southern blot analysis be used to confirm the identity of specific product bands. Specific product can also be identified using a diagnostic restriction endonuclease digestion if the amplified cDNA sequence contains a known restriction site.

6. Analyze 10 to 20 µl of the amplified sample, using agarose gel electrophoresis and ethidium bromide staining, and the appropriate molecular size standards. If the positive control RNA was used, a single 720-bp band will be visible (see Section 4.2, *Testing the 3′ RACE System*).

Seikagaku Corporation Exhibit 2004 - Page 163 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

4 Troubleshooting Guide

4.1 General Suggestions

Problem	Possible Cause	Suggested Remedy
No bands after electrophoretic analysis of amplified products	Procedural error in first strand cDNA synthesis or PCR	Use the control RNA to verify the efficiency of the first strand reaction (see Section 4.2, *Testing the 3´ RACE System).*
	RNase contamination	Add the control RNA to the sample to determine if RNase is present in the first strand reaction. Maintain aseptic conditions to prevent RNase contamination (see Section 4.3, *Minimizing RNase Contamination*). Use RNase inhibitor during first strand cDNA synthesis.
	Inhibitors of RT present	Remove inhibitors by ethanol precipitation of the mRNA preparation before the first strand reaction. Include a 70% (v/v) ethanol wash of the mRNA pellet. **Notes:** Inhibitors of RT include sodium dodecyl sulfate (SDS), EDTA, guanidinium salts, glycerol (>35%), sodium pyrophosphate, and spermidine (18). Test for the presence of inhibitors by mixing 1 µg of sample RNA ± control RNA and compare yields of first strand cDNA or PCR product.
	Polysaccharide and small RNA (tRNA and 5SRNA) coprecipitation with mRNA	Ethanol-precipitate the RNA preparation; treat the pellet as described in Section 5.4, *Lithium Chloride Purification of RNA Preparation* (17).
	Target mRNA contains strong transcriptional pauses	Maintain an elevated temperature after the annealing step and increase the temperature of first strand reaction (up to 50°C) (see Section 5.1, *Alternative Protocol for First Strand cDNA Synthesis of Transcripts with High GC Content*).
	Too much first strand reaction was used in the PCR	Dilute cDNA reaction 10- to 100-fold. Use no more than 10% of the first strand product in PCR.
	Polymerase from an archaeobacteria was used with dUMP primers	Use AUAP and non-dUMP GSP in PCR using Elongase® or an archaeobacterial polymerase. Use *Taq* DNA polymerase for PCR with the UAP and dUMP containing GSP.
Unexpected bands after electrophoretic analysis of "nested" amplification products	Contamination by genomic DNA	To test if products were derived from genomic DNA, perform first strand reaction without SuperScript™ II RT. Pretreat RNA as described in Section 5.3, *DNase I Digestion of RNA Preparation.*
	Spurious priming in the PCR	Vary the parameters of the PCR according to *Taq* DNA polymerase instructions and/or use hot start for PCR. Incorporate a preamplification heating step (44,45).

Seikagaku Corporation Exhibit 2004 - Page 164 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Problem	Possible Cause	Suggested Remedy
Poor cloning efficiency	Inefficient ligation	Increase the incubation time in the ligation reaction; decrease the temperature.
		Ensure the removal of dNTPs prior to ligation by chromatographic separation.
		Ensure the removal of vector stuffer fragments by chromatographic separation.
	Poor restriction endonuclease digestion due to residual bound *Taq* DNA polymerase	Treat the PCR products with proteinase K (47) (see Section 5.8, *Proteinase K Treatment of 3´ RACE Products).*
	Restriction endonuclease does not digest at the end of molecule	Add nucleotides to 5´ end of primer. Use a different restriction endonuclease.

4.2 Testing the 3´ RACE System

The control RNA provided with the 3´ RACE System can be used to trouble-shoot both the first strand reaction and the amplification reaction. Use the following protocols, as needed, to troubleshoot for particular problems.

To perform the 3´ RACE procedure using the control RNA, you may need the following items, *in addition to* those listed in Section 3.2, *Advance Preparations*.

- [α-^{32}P]dCTP (3,000 Ci/mmol; 10 mCi/ml)
- glass-fiber filters
- heat lamp
- 10% (w/v) TCA containing 1% (w/v) sodium pyrophosphate
- 95% ethanol

4.2.1 First Strand cDNA Synthesis

1. Label two autoclaved 0.5-ml microcentrifuge tubes "A" and "B." Tube A will have an addition of radioisotope to determine the efficiency of first strand synthesis. An aliquot from tube B will be used for amplification.
2. Prepare the RNA:primer mixture in a sterile 0.5-ml tube:

Note: Up to 1 µg of your total RNA preparation may be added to this mixture in order to assess the effect of this RNA on first strand synthesis and subsequent amplification.

	Volume (µl)	
Component	Tube A	Tube B
control RNA (50 ng/µl)	2	2
AP (10 µM)	1	1
DEPC-treated water	8	9
final volume	11	12

3. Incubate at 70°C for 10 min and place on ice for 1 min. Collect the contents of each tube by brief centrifugation and add the following to each tube:

	Volume (µl)	
Component	Tube A	Tube B
10X PCR buffer	2	2
25 mM $MgCl_2$	2	2
10 mM dNTP mix	1	1
0.1 M DTT	2	2
SuperScript™ II RT (200 units/µl)	1	1

Seikagaku Corporation Exhibit 2004 - Page 165 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

4. Add 1 µl of [α-^{32}P]dCTP (3,000 Ci/mmol; 10 mCi/ml) to tube A. Final volume will be 20 µl.
5. Mix gently and collect the reaction by brief centrifugation.
6. Incubate for 42°C for 50 min.
7. Terminate both reactions by incubating at 70°C for 15 min. Place on ice for 10 min.
8. Collect each reaction by brief centrifugation. Add 1 µl of RNase H to tube B and incubate for 20 min at 37°C. Proceed with tube B to Section 4.2.2, *Amplification of the Control-Synthesized cDNA.*
9. Add 80 µl of distilled water to tube A and mix gently.
10. Remove two 5-µl aliquots from tube A and spot the aliquots onto glass-fiber filters. Dry one of the filters under a heat lamp or at room temperature. This filter will be used to determine the specific activity of the dCTP reaction.
11. Wash the other filter three times in sequence, for 5 min each time, with 50 ml of ice-cold, 10% (w/v) TCA containing 1% (w/v) sodium pyrophosphate. Wash the filter once with 50 ml of 95% ethanol at room temperature for 2 min. Dry the filter under a heat lamp or at room temperature. This filter will be used to determine the yield of first strand cDNA.
12. Count both filters in standard scintillant to determine the amount of ^{32}P in the reaction, as well as the amount of ^{32}P that was incorporated.
13. Using equation 1, determine the specific activity (SA) of the dCTP in the first strand reaction from the counts obtained from the *unwashed* filter:

$$\text{SA (cpm/pmol dCTP)} = \frac{\text{cpm/5 µl}}{\text{500 pmol dCTP/5 µl}}$$

14. Using equation 2, determine the yield of cDNA from the counts obtained from the *washed* filter and the specific activity calculated from the unwashed filter:

$$\text{Amount of cDNA (µg)} = \frac{\text{(cpm)} \times \text{(100 µl/5 µl)} \times \text{(4 pmol dNTP/pmol dCTP)}}{\text{(SA)} \times \text{(3,030 pmol dNTP/µg cDNA)}}$$

The yield calculated for the labeled cDNA in tube A can be assumed as equivalent to that of the unlabeled cDNA in tube B.

15. Following first strand cDNA synthesis from the control RNA, you may wish to analyze the remaining cDNA in tube A by alkaline agarose gel electrophoresis or denaturing polyacrylamide gel electrophoresis (PAGE).

Note: The first strand yield using the control RNA should be 20% to 50%.

Note: Electrophoretic analysis should yield a prominent 891-bp band representing 25% to 50% of the cDNA synthesized.

4.2.2 Amplification of the Control-Synthesized cDNA

Following first strand cDNA synthesis from the control RNA, the control GSP (figure 3) and the UAP can be used to troubleshoot the amplification reaction. Electrophoretic analysis of DNA products amplified using the control primers should yield a prominent 720-bp band.

5′–CAU CAU CAU CAU GAC CGT TCA GCT GGA TAT TAC–3′

Figure 3. Sequence of the control GSP.

Seikagaku Corporation Exhibit 2004 - Page 166 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

1. In each of three fresh 0.5-ml microcentrifuge tubes, prepare a PCR mixture as in step 1 of Protocol 2, substituting 1 µl of the control GSP for the user-defined GSP and using the UAP amplification primer. Place tubes on ice.
2. In separate 1.5-ml microcentrifuge tubes, prepare serial dilutions of the control cDNA from tube B in sterile, distilled water. Dilute 1:100, 1:1,000, 1:10,000, 1:100,000, and 1:1,000,000. Use the three highest dilutions:

   control 1: $1:10^4$ ($\sim 10^6$ input RNA molecules/µl)
   control 2: $1:10^5$ ($\sim 10^5$ input RNA molecules/µl)
   control 3: $1:10^6$ ($\sim 10^4$ input RNA molecules/µl)

3. Add 2 µl from each control dilution to a separate reaction tube from step 1. Mix and layer 75 µl of mineral oil over each reaction. Collect each reaction by brief centrifugation.

Note: Refer to Protocol 2, Section 3.4, for additional information.

4. Perform steps 3 through 6 in Protocol 2. Amplify with 30 cycles of 94°C for 45 s, 55°C for 25 s, and 72°C for 3 min. Following agarose gel electrophoresis and ethidium bromide staining, you should be able to visualize a 720-bp band for all three concentrations of input cDNA (see figure 4).

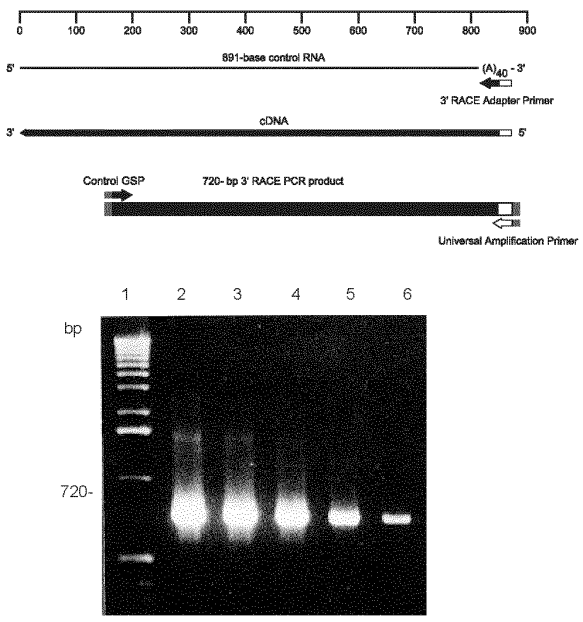


Figure 4. Agarose gel analysis of 3′ RACE products using control RNA. A 100-ng sample of control RNA was used for first strand cDNA synthesis as described in this section. The resulting cDNA was diluted 10^3 to 10^7 fold. 2 µl of each dilution was amplified according to Protocol 2 (Section 3.4) and one-tenth (5 µl) of the reaction product was analyzed by 1% agarose gel electrophoresis in the presence of 0.5 µg/ml of ethidium bromide. Lane 1: 1 Kb DNA Ladder. Lane 2: 2×10^7 input RNA molecules per reaction. Lane 3: 2×10^6 input RNA molecules per reaction. Lane 4: 2×10^5 input RNA molecules per reaction. Lane 5: 2×10^4 input RNA molecules per reaction. Lane 6: 2×10^3 input RNA molecules per reaction.

Seikagaku Corporation Exhibit 2004 - Page 167 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

4.3 Minimizing RNase Contamination

Successful cDNA synthesis demands an RNase-free environment at all times, which will generally require the same level of care used to maintain aseptic conditions when working with microorganisms. Several additional guidelines should be followed:

1. Never assume that anything is RNase-free. Use sterile pipets, centrifuge tubes, culture tubes, or any similar labware that is explicitly stated to be sterile. **Note:** Wear latex gloves for all manipulations involving RNA.
2. Dedicate a separate set of automatic pipets for manipulating RNA and the buffers and enzymes used to synthesize cDNA.
3. Avoid using any recycled glassware unless it has been specifically rendered RNase-free by rinsing with 0.5 N NaOH followed by copious amounts of sterilized, distilled water. Alternatively, bake glassware at 150°C for 4 h.
4. Microcentrifuge tubes can generally be taken from an unopened box, autoclaved, and used for all cDNA work. RNase-free microcentrifuge tubes can be purchased from several suppliers. If it is necessary to decontaminate untreated tubes, soak the tubes overnight in a 0.01% (v/v) aqueous solution of diethylpyrocarbonate (DEPC), rinse with sterilized, distilled water, and autoclave.
5. If made with RNase-free labware, most solutions can be made from reagent-grade materials and distilled water, and then autoclaved. Prepare heat-sensitive solutions using sterilized, distilled water, and filtering through a 0.2 µm using sterilized, disposable unit.
6. Most aqueous buffer solutions can be treated with 0.01% (v/v) DEPC and autoclaved. **Note:** Buffers containing primary amines (such as tris) cannot be effectively treated by this method.
7. Use aerosol-resistant pipet tips.

Seikagaku Corporation Exhibit 2004 - Page 168 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Additional Protocols

5

5.1 First Strand cDNA Synthesis of Transcripts with High GC Content

Problems with 3´ RACE due to secondary structure of the target mRNA may be overcome by increasing the volume and temperature of the RT reaction. To avoid secondary RNA structure, shift the primer/RNA mix directly from 70°C to 50°C and prewarm the complete 2X reaction mix to 42°C before adding it to the primer and RNA. Use of a thermocycler simplifies the multiple temperature shifts in RT PCR and can help to prevent formation of secondary structure in RNA.

First Strand Synthesis Conditions
- 20 mM Tris-HCl (pH 8.4 at 22°C)
- 50 mM KCl
- 2.5 mM $MgCl_2$
- 10 mM DTT
- 500 µM each dATP, dCTP, dGTP, and dTTP
- 500 nM AP
- ≤1 µg total RNA
- 200 units SuperScript™ II RT
- Reaction volume: 50 µl

1. Mix and briefly centrifuge each component before use.
2. Combine up to 1 µg of total or 50 ng of poly(A)-selected RNA and DEPC-treated water to a final volume of 24 µl in a 0.5-ml microcentrifuge tube.
3. Add 1 µl of the 10 µM AP solution, mix gently, and collect by brief centrifugation.
4. Heat each sample to 70°C for 10 min and immediately transfer to 50°C.
5. In a separate tube, assemble the following reaction mixture, adding each component in the indicated order:

Component	Volume (µl)
DEPC-treated water	7.5
10X PCR buffer	5
25 mM $MgCl_2$	5
10 mM dNTP mix	2.5
0.1 M DTT	5

6. Prewarm the reaction mix to 42°C.
7. Add the prewarmed reaction mixture and 1 µl (200 units) of SuperScript™ II RT to the RNA/primer mixture, mix gently, and continue to incubate at 50°C for 50 min.
8. Terminate the reaction by incubating the tubes at 70°C for 15 min. Chill on ice.
9. Collect the reaction by brief centrifugation. Add 1 µl of RNase H and incubate for 20 min at 37°C before proceeding to Protocol 2 (see Section 3.4).

Seikagaku Corporation Exhibit 2004 - Page 169 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Additional Protocols

5.2 TRIzol® Reagent Isolation of Total RNA

TRIzol® Reagent is a ready-to-use reagent for the isolation of total RNA from cells and tissues. The reagent, a monophasic solution of phenol and guanidine isothiocyanate, is an improvement to the single-step RNA isolation method developed by Chomczynski and Sacchi (19,20). During sample homogenization or lysis, TRIzol® Reagent maintains the integrity of the RNA, while disrupting cells and dissolving cell components. Addition of chloroform followed by centrifugation, separates the solution into an aqueous phase and an organic phase. RNA remains exclusively in the aqueous phase. The RNA is recovered from the aqueous phase by precipitation with isopropanol. The basic protocol is below. More details are found on the product information provided with the TRIzol® Reagent.

Note: Unless otherwise stated, the procedure is carried out at room temperature, and reagents are at room temperature.

Materials Required

- TRIzol® Reagent,
- chloroform,
- isopropanol,
- DEPC-treated water, and
- 75% ethanol (prepared using DEPC-treated water).

1. HOMOGENIZATION (see notes 1 and 2)
   a. *Tissues*

      Homogenize tissue samples in 1 ml of TRIzol® Reagent per 50 to 100 mg of tissue using a glass-Teflon® or power homogenizer. The sample volume should not exceed 10% of the volume of TRIzol® Reagent used for homogenization.

   b. *Cells Grown in Monolayer*

      Lyse cells directly in a culture dish by adding 1 ml of TRIzol® Reagent to a 3.5-cm diameter dish, and passing the cell lysate several times through a pipette. The amount of TRIzol® Reagent added is based on the area of the culture dish (1 ml per 10 cm^2).

   c. *Cells Grown in Suspension*

      Pellet cells by centrifugation. Lyse cells in TRIzol® Reagent by repetitive pipetting. Use 1 ml of the reagent per 5×10^6 to 10×10^6 of animal, plant or yeast cells, or per 1×10^7 bacterial cells. Washing cells before addition of TRIzol® Reagent should be avoided to minimize mRNA degradation. Disruption of yeast and bacterial cells may require a homogenizer.

2. Incubate the homogenized samples for 5 min at room temperature. Add 0.2 ml of chloroform per 1 ml of TRIzol® Reagent. Cap sample tubes securely. Shake tubes vigorously by hand for 15 s and incubate for 2 to 3 min. Centrifuge the samples at no more than 12,000 × *g* for 15 min at 4°C. Following centrifugation, the mixture separates into a lower red, phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. The volume of the aqueous phase is about 60% of the volume of TRIzol® Reagent used for homogenization.
3. Transfer the aqueous phase to a fresh tube. Add 0.5 ml of isopropanol per 1 ml of TRIzol® Reagent used for the initial homogenization. Incubate samples at room temperature for 10 min and centrifuge at no more than 12,000 × *g* for 10 min.

Seikagaku Corporation Exhibit 2004 - Page 170 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

4. Remove the supernate. Resuspend the RNA pellet once in 75% ethanol, adding at least 1 ml of 75% ethanol per 1 ml of TRIzol® Reagent used for the initial homogenization. Mix the sample by vortexing and centrifuge at no more than 7,500 × *g* for 5 min at 4°C. The RNA precipitate can be stored in 75% ethanol at 4°C for at least one week, or at least one year at –20°C.

Note: Use only DEPC-treated water to resuspend RNA.

5. Briefly dry the RNA pellet (air-dry or vacuum-dry for 5 to 10 min). Do not dry the RNA by centrifugation under vacuum. It is important not to let the RNA pellet dry completely as this will greatly decrease its solubility. Partially dissolved RNA samples have an $A_{260}/_{280}$ ratio <1.6. Dissolve RNA in DEPC-treated water and incubate for 10 min at 55°C to 60°C.

Notes:

1. To facilitate isolation of RNA from small quantities of sample (10^6 cells or <10 mg tissue) perform homogenization (or lysis) of samples in 0.8 ml of TRIzol® Reagent, and add 5 to 10 µg of glycogen as carrier.
2. An additional isolation step may be required for samples with a high content of proteins, fat, polysaccharides or extracellular material such as muscles, fat tissue, and tuberous parts of plants. Following homogenization, remove insoluble material from the homogenate by centrifugation at 12,000 × *g* for 10 min at 4°C. The resulting pellet contains extracellular membranes, polysaccharides, and high molecular weight DNA, while the supernate contains RNA. In samples from fat tissue, an excess of fat collects as a top layer which should be removed. Transfer the clear supernate to a fresh tube and proceed with the phase separation.

5.3 DNase I Digestion of RNA Preparation

If amplification products are detected from the PCR in the absence of SuperScript™ II RT, it may be necessary to eliminate residual genomic DNA from the RNA sample. After confirming the efficiency of the first strand synthesis reaction with the control RNA, use the following protocol to remove genomic DNA from the total RNA preparation.

1. Add the following to a 0.5-ml microcentifuge tube on ice:

Component	Amount
total RNA	1–2 µg
10X reaction buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl, 20 mM $MgCl_2$]	1 µl
Amplification grade DNase I (1 unit/µl)	1 µl
DEPC-treated water	to 10 µl

Note: The procedure requires careful pipetting of all solutions so that the concentration of divalent metal cation (Mg^{2+} and Ca^{2+}) is precise. Because the DNase I must be heated to 65°C to inactivate the enzyme, the concentration of free divalent metal ion must be low enough (<1 mM) after the addition of the EDTA to prevent chemical hydrolysis of the RNA.

2. Incubate at room temperature for 15 min.
3. Add 1 µl of 25 mM EDTA.
4. Incubate for 15 min at 65°C to heat inactivate the DNase I, then place on ice for 1 min. Collect the reaction by brief centrifugation. This mixture can be used directly for reverse transcription.

5.4 Lithium Chloride Purification of RNA Preparation

If you obtained poor first strand cDNA synthesis from your RNA preparation, the RNA may have coprecipitated with polysaccharides or small RNAs (tRNA and 5SRNA) during the RNA isolation procedure. You may be able to recover highly purified mRNA using the following protocol adapted from Sambrook *et al.* (17).

1. To the RNA sample in DEPC-treated water, add 0.1 volume 8 M LiCl (RNase-free) and vortex. Incubate on ice for 2 h.

Seikagaku Corporation Exhibit 2004 - Page 171 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

2. Centrifuge at 14,000 × *g* for 30 min at 4°C.
3. Remove the supernate, being careful not to disturb the pellet (the pellet may be difficult to see). Dissolve the pellet in 200 µl of DEPC-treated water by drawing the pellet in and out of a sterilized pipet tip.
4. Repeat steps 1 through 3.
5. Add 0.1 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of absolute ethanol (–20°C). Place the tube at –20°C for 30 min. Centrifuge at 14,000 × *g* for 30 min at 4°C.
6. Remove the supernate carefully. Overlay the pellet with 100 µl of 70% ethanol (–20°C) and centrifuge at 14,000 × *g* for 10 min at 4°C. Remove the supernate and air dry the RNA pellet at room temperature.
7. Dissolve the pellet in 10 to 100 µl of DEPC-treated water.
8. If your starting cell or tissue sample was small (10^3 to 10^5 cells), dissolve the pellet in 13 µl of DEPC-treated water and use this entire amount for first strand synthesis. If your starting sample was large (10^6 to 10^7 cells), spectrophotometrically determine the concentration of total RNA recovered in step 5 before proceeding to Protocol 1 (Section 3.3).

5.5 Nested Amplification from an Agarose Plug

This procedure is intended to provide added specificity to the amplification procedure prior to cloning the 3′ RACE products. The following protocol should be performed after completion of Protocol 2 (agarose gel electrophoretic analysis and RACE product visualization by ethidium bromide staining), but prior to performing the cloning procedures described in this chapter.

Note: Sequences specific for downstream cloning manipulations (see Section 2.6, *Design of Gene-Specific Primer*) must be designed into GSP2, not GSP1.

1. Using a capillary pipet, remove a small (up to 5 µl) agarose plug from the area of the gel containing the band of interest.
2. Prepare an amplification reaction mix as in step 1 of Protocol 2, *except* use a second GSP ("GSP2") in place of the original GSP.
3. Perform steps 2 through 7 in Protocol 2 and proceed to subsequent cloning procedures.

Seikagaku Corporation Exhibit 2004 - Page 172 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

5.6 Cloning Using T4 DNA Polymerase

1. Following amplification, remove dNTPs from the reaction by agarose gel electrophoresis. Resuspend 200 to 500 ng of the reaction in 34 µl of sterilized, distilled water.
2. To the purified amplification product, add the following:

Component	Volume (µl)
10X Tris-acetate buffer [0.33 M Tris-acetate (pH 8.0), 0.66 M potassium acetate, 0.1 M magnesium acetate, 5 mM DTT, 1 mg/ml BSA]	5
dTTP (1 mM)	5
dATP (1 mM)	5
T4 DNA polymerase (1 unit/µl)	1

3. Incubate the reaction at 37°C for 15 min. Heat inactivate the enzyme by incubating at 65°C for 10 min. Ethanol precipitate the reaction (17) and resuspend in 5 µl of TE Buffer.
4. Prepare pSPORT 1 or other appropriate vector for ligation by digestion with *Acc* I and *Not* I (17). Purify the vector from the excised fragment by agarose gel electrophoresis.
5. Combine ~100 ng of nondephosphorylated *Not* I-*Acc* I-cut vector with 500 ng of the T4 DNA polymerase-treated 3′ RACE product and appropriate ligation reaction buffer to a final volume of 20 µl. Add 1 or 2 units of T4 DNA ligase and incubate at 20°C to 25°C for 2 to 4 h.
6. Dilute the ligation reaction 1:10 in TE [10 mM Tris-HCl (pH 7.5), 1 mM EDTA]. Transform competent cells such as MAX Efficiency® or Library Efficiency® DH5α™ Competent Cells using 1 to 2 µl of the vector-ligated DNA.

Note: For additional information about ligation reactions using T4 DNA ligase, see reference 48 or further information on bacterial transformation (49).

5.7 Proteinase K Treatment of 3′ RACE Products

If you obtained poor cloning efficiency following amplification, residual bound *Taq* DNA polymerase may have hindered restriction endonuclease digestion, resulting in poor vector/insert ligation. Use the following protocol to ensure complete removal of *Taq* DNA polymerase after amplification (47).

1. Take sample from step 5 of protocol 2 (see Section 3.4) and add 0.1 volume of proteinase K solution [500 µg/ml proteinase K, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA].
2. Incubate for 30 min at 37°C and then for 10 min at 65°C.
3. Extract once with phenol:chloroform (1:1) and once with chloroform.
4. Ethanol-precipitate by addition of 0.1 volume of sodium acetate (pH 5.2) and 2 volumes of absolute ethanol (–20°C). Place tube at –20°C for 30 min.
5. Centrifuge at 13,000 × *g* for 30 min at 4°C. Remove supernate, rinse pellet with 70% ethanol, and air-dry the pellet at room temperature.
6. Resuspend the sample in 10 µl of TE buffer [10 mM Tris (pH 8.0), 1 mM EDTA] before advancing to cloning procedures.

Seikagaku Corporation Exhibit 2004 - Page 173 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

6 References

1. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA 85,* 8998.
2. Ohara, O., Dorit, R.L., and Gilbert, W. (1989) *Proc. Natl. Acad. Sci. USA 86,* 5673.
3. Loh, E.Y., Elliott, J.F., Cwirla, S., Lanier, L.L., and Davis, M.M. (1989) *Science 243,* 217.
4. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985) *Science 230,* 1350.
5. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) *Science 239,* 487.
6. Loh, E.Y. (1991) *Methods 2,* 11.
7. Berchtold, M.W. (1989) *Nucleic Acids Res. 17,* 453.
8. Brow, M.A.D. (1990) *PCR Protocols: A Guide to Methods and Applications* (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds.) p. 189, Academic Press, San Diego.
9. Adams, S.M. and Blakesley, R. (1991) *Focus® 13,* 56.
10. Frohman, M.A. and Martin, G.R. (1989) *Techniques 1,* 165.
11. Frohman, M.A. (1990) *PCR Protocols: A Guide to Methods and Applications* (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds.) p. 28, Academic Press, San Diego.
12. Buckler, A.J., Chang, D.D., Graw, S.L., Brook, D., Haber, D.A., Sharp. P.A., and Housman, D.E. (1991) *Proc. Natl. Acad. Sci. USA 88,* 4005.
13. Nisson, P.E., Rashtchian, A., and Watkins, P.C. (1991) *PCR Meth. and Appl. 1,* 120.
14. Buchman, G.W. and Rashtchian, A. (1992) *Focus® 14,* 2.
15. Buchman, G.W., Schuster, D.M., and Rashtchian, A. (1992) *Focus® 14,* 41.
16. Rashtchian, A., Buchman, G.W., Schuster, D.M., and Berninger, M.S. (1992) *Anal. Biochem. 206,* 91.
17. Sambrook J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
18. Gerard, G. F. (1994) *Focus® 16,* 102.
19. Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem. 162,* 156.
20. Chomczynski, P. (1993) *Biotechniques 15,* 532.
21. Simms, D., Cizdziel, P. E., and Chomczynski, P. (1993) *Focus® 15,* 99.
22. Simms, D. (1993) *Focus 15,* 6.
23. Gerard, G.F., D'Alessio, J.M., and Kotewicz, M.L. (1989) *Focus® 11,* 66.
24. D'Alessio, J.M., Gruber, C.E., Cain, C., and Noon, M.C. (1990) *Focus® 12,* 47.
25. Gerard, G. F., Schmidt, B. J., Kotewicz, M. L. and Campbell, J. H. (1992) *Focus® 14,* 91.
26. Hu, A.W., D'Alessio, J.M., Gerard, G.F., and Kullman, J. (1990) *Focus® 13,* 26.
27. Scharf, S.J. (1990) *PCR Protocols: A Guide to Methods and Applications* (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds.) p. 84, Academic Press, San Diego.
28. Stoker, A. W. (1990) *Nucleic Acids Res. 18,* 4290.
29. Saiki, R.K. (1989) *PCR Technology:. Principles and Applications for DNA Amplification* (Erlich, H.A., ed.) pps. 17–22, Stockton Press, New York.

Seikagaku Corporation Exhibit 2004 - Page 174 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

30. Saiki, R.K. (1990) *PCR Protocols: A Guide to Methods and Applications* (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. eds.) pps. 13–20, Academic Press, San Diego.
31. Rychlik, W. (1995) Biotechniques 18,84.
32. Rychlik, W. and Rhodes, R.E. (1989) *Nucleic Acids Res. 17,* 8543.
33. Rychlik, W., Spencer, W.J., and Rhodes, R.E. (1990) *Nucleic Acids Res. 18,* 6409.
34. Lowe, T., Sharefkin, J., Yang, S.Q., and Diffenbach, C.W. (1990) *Nucleic Acids Res. 18,* 1757.
35. Hiller, L. and Green, P. (1991) *PCR Meth. and Appl. 1,* 124.
36. Craxton, M. (1991) *Methods 3,* 20.
37. Bhat, G.J., Lodes, M.J., Myler, J., and Stuart, K.D. (1991) *Nucleic Acids Res. 19,* 398.
38. Shuldiner, A.R., Scott, L.A., and Roth, J. (1990) *Nucleic Acids Res. 18,* 1920.
39. Duncan, B.K. (1981) in *The Enzymes* (Boyer, P., ed.), 3rd Ed., Part A, p. 565, Academic Press, New York.
40. Friedberg, E.C., Bonura, T., Radany, E.H., and Love, J.D. (1981) in *The Enzymes* (Boyer, P., ed.), 3rd Ed., Part A, p. 251, Academic Press, New York.
41. Eckert, K.A. and Kunkel, T.A. (1991) *PCR Meth. and Appl. 1,* 17.
42. Kwok , S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C., and Sninsky, J.J. (1990) *Nucleic Acids Res. 18,* 999.
43. Innis, M.A. and Gelfand, D.H. (1990) *PCR Protocols: A Guide to Methods and Applications* (Innis, M.A., Gelfand, D. H., Sninsky, J.J., and White, T.J., eds) pps. 3–12, Academic Press, San Diego.
44. D'Aquila, R.T., Bechtel, L.J., Videler, J.A., Eron, J.J., Gorczyca, P., and Kaplan, J.C. (1991) *Nucleic Acids Res. 19,* 3749.
45. Mullis, K.B. (1991) *PCR Meth. and Appl. 1,* 1.
46. Gustafson, C.E., Alm, R.A., and Trust, T.J. (1993) *Gene 123,* 241.
47. Crowe, J.S., Cooper, H.J., Smith, M.A., Sims, M.J., Parker, D., and Gewert, D. (1991) *Nucleic Acids Res. 19,* 184.
48. Life Technologies, Inc. (1996) *Focus on Applications: T4 DNA Ligase (Technical Bulletin 15244-2).*
49. Life Technologies, Inc., A Division of Invitrogen Corp. (2001) *Guide to Bacterial Transformation.*

Seikagaku Corporation Exhibit 2004 - Page 175 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

7 Related Products

Product	Size	Cat. No.
Primers:		
Custom Primers		Please inquire
Abridged Universal Amplification Primer	3.1 µg	10541-019
Universal Amplification Primer	5 µg	18382-010
5´ RACE Abridged Anchor Primer	5.7 µg	10630-010
5´ RACE Anchor Primer	7.5 µg	18388-017
3´ RACE Adapter Primer	2.3 µg	10542-017
Products for Transformation and Cloning:		
MAX Efficiency® DH5α™ Competent Cells	1 ml (5 × 0.2 ml)	18258-012
MAX Efficiency® DH10B™ Competent Cells	1 ml (5 × 0.2 ml)	18297-010
Library Efficiency® DH5α™ Competent Cells	1 ml (5 × 0.2 ml)	18263-012
MAX Efficiency® STBL2™ Competent Cells	1 ml (5 × 0.2 ml)	10268-019
S.O.C. Medium	10 × 10 ml	15544-034
Enzymes for Amplification:		
Taq DNA Polymerase, Native	100 units 500 units 1,500 units (3 × 500 units)	18038-018 18038-042 18038-067
Taq DNA Polymerase, Recombinant	100 units 500 units 1,500 units (3 × 500 units)	10342-053 10342-020 10342-046
Platinum® *Taq* DNA Polymerase	250 units	10966-026
Elongase® Amplification System	100 reactions	10481-018
Elongase® Enzyme Mix	100 reactions 500 reactions	10480-010 10480-028
Other Modifying Enzymes:		
T4 DNA Ligase	100 units 500 units 2,000 units (4 × 500 units)	15224-017 15224-025 15224-090
T4 DNA Ligase High Concentration	250 units	15224-041
DNase I, Amplification Grade	100 units	18068-015
T4 DNA Polymerase	50 units 250 units	18005-017 18005-025

Seikagaku Corporation Exhibit 2004 - Page 176 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Product	Size	Cat. No.
Ribonuclease H	30 units	18021-014
	120 units	18021-071
SuperScript™ II Reverse Transcriptase	2,000 units	18064-022
	10,000 units	18064-014
	40,000 units (4 × 10,000 units)	18064-071
RNaseOUT™, Recombinant RNase Inhibitor	5,000 units	10777-019
Other Related Products:		
TRIzol® Reagent	100 ml	15596-026
	200 ml	15596-018
5´ RACE System for Rapid Amplification of cDNA Ends, Version 2.0	10 reactions	18374-058
100 bp DNA Ladder	50 µg	15628-019
	250 µg	15628-050
1 Kb DNA Ladder	250 µg	15615-016
	1,000 µg	15615-024
DEPC-Treated Water	4 × 1.25 ml	10813-012
GlassMAX RNA Isolation Spin Cartridge System	50 reactions	18385-039

Seikagaku Corporation Exhibit 2004 - Page 177 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023



United States Headquarters:
Invitrogen Corporation
1600 Faraday Avenue
Carlsbad, California 92008
Tel: 1 760 603 7200
Tel (Toll Free): 1 800 955 6288
Fax: 1 760 603 7229
Email: tech_service@invitrogen.com

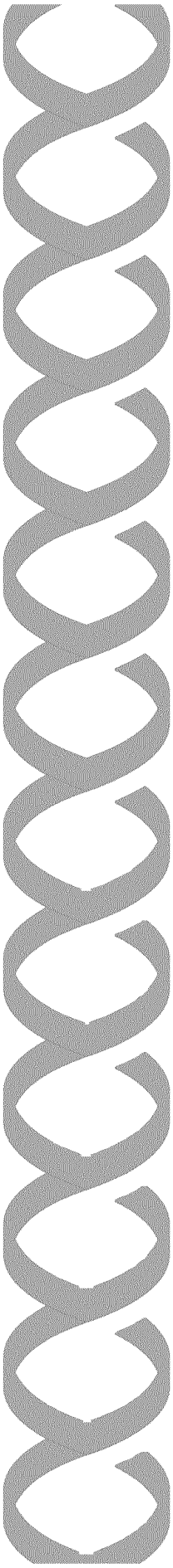
European Headquarters:
Invitrogen Ltd
3 Fountain Drive
Inchinnan Business Park
Paisley PA4 9RF, UK
Tel (Free Phone Orders): 0800 269 210
Tel (General Enquiries): 0800 5345 5345
Fax: +44 (0) 141 814 6287
Email: eurotech@invitrogen.com

International Offices:
Argentina 5411 4556 0844
Australia 1 800 331 627
Austria 0800 20 1087
Belgium 0800 14894
Brazil 0800 11 0575
Canada 800 263 6236
China 10 6849 2578
Denmark 80 30 17 40

France 0800 23 20 79
Germany 0800 083 0902
Hong Kong 2407 8450
India 11 577 3282
Italy 02 98 22 201
Japan 03 3663 7974
The Netherlands 0800 099 3310
New Zealand 0800 600 200
Norway 00800 5456 5456

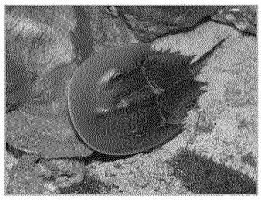
Spain & Portugal 900 181 461
Sweden 020 26 34 52
Switzerland 0800 848 800
Taiwan 2 2651 6156
UK 0800 838 380
For other countries see our website

www.invitrogen.com



An official website of the United States government
Here's how you know

Genome assembly Limulus_polyphemus-2.1.2 reference



Download datasets curl

NCBI RefSeq assembly
GCF_000517525.1

Submitted GenBank assembly
GCA_000517525.1

Taxon
Limulus polyphemus (Atlantic horseshoe crab)

WGS project
AZTN01

Assembly type
haploid

Submitter
Washington University (WashU)

Date
Jan 3, 2014

View the legacy Assembly page

Feedback

View annotated genes

See in Genome Data Viewer

BLAST the reference genome

Assembly statistics

	RefSeq	GenBank
Genome size	1.8 Gb	1.8 Gb
Total ungapped length	1.7 Gb	1.7 Gb

	RefSeq	GenBank
Number of organelles	1	0
Number of scaffolds	286,792	286,792
Scaffold N50	254.1 kb	254.1 kb
Scaffold L50	1,712	1,712
Number of contigs	469,509	469,509
Contig N50	11.4 kb	11.4 kb
Contig L50	41,763	41,763
GC percent	33.5	33.5
Genome coverage	18.0x	18.0x
Assembly level	Scaffold	Scaffold

Sample details

BioSample ID
SAMN02953878

Description
Sample from Limulus polyphemus

Owner name
Washington University School of Medicine, The Genome Institute

Sex
male

Tissue
muscle

Geo loc name
USA: Woods Hole, MA

Collection date
2008-01

Assembly methods

Sequencing technology
Roche 454

Comment
Background:
The Limulus polyphemus DNA used for sequencing was derived from a single male animal (carapace length approximately 16 cm) collected January 2008 in Woods Hole, MA. Muscle tissue was extracted from the limbs, and genomic DNA was prepared in the lab of Nipam Patel by Meredith Protas.
Total sequence coverage of Roche 454 reads was approximately 18C (fragments, 3 and 8kbs) using a genome size estimate of 1.5Gb. The combined sequence reads were assembled using the Newbler software (Roche). This draft assembly was referred to as Limulus polyphemus 2.1.1. This version has been gap filled, error corrected with
12X Illumina reads, and cleaned of contaminating contigs. The assembly is made up of a total of 286,840 scaffolds with an N50 scaffold length of 238kb (N50 contig length was 11.4kb). The assembly spans 1.7Gb.
This work was funded by NIH grant HG003079 to RKW, The Genome Institute at Washington University School of Medicine.
For questions regarding this Limulus polyphemus 2.1.1 assembly please contact Patrick Minx, Washington University School of Medicine (pminx@genome.wustl.edu). Funding for the sequence characterization of the Limulus polyphemus genome was provided by the Office of Director, National Institutes of Health (NIH). **********
Sequence and Assembly Credits:
Sequencing DNA source credit: Nipam Patel and Meredith Protas
DNA source can be obtained from: Nipam Patel Dept. of Molecular and Cell Biology 519A LSA 3200 UC Berkeley Berkeley, CA 94720-3200

Genome Sequence - The Genome Institute, Washington University School of Medicine
Sequence Assembly - The Genome Institute, Washington University School of Medicine
Citation upon use of this assembly in a manuscript:
It is requested that users of this Limulus polyphemus 2.1.1 sequence assembly acknowledge Richard K. Wilson and The Genome Institute, Washington University School of Medicine in any publications that result from use of this sequence assembly.
Assembly Stats
*** Contiguity: Contig *** Total contig number: 469557 Total contig bases: 1705828287 bp Average contig length: 3633 bp Maximum contig length: 133547 bp N50 contig length: 11441 bp N50 contig number: 41760
*** Contiguity: Supercontig *** Total supercontig number: 286840 Average supercontig length: 5947 bp Maximum supercontig length: 4914972 bp N50 supercontig length: 238003 bp N50 supercontig number: 1693
*** Scaffold Distribution *** Scaffolds > 1M: 128 Scaffold 250K--1M: 1451 Scaffold 100K--250K: 2309 Scaffold 10--100K: 8850 Scaffold 5--10K: 3013 Scaffold 2--5K: 8455 Scaffold 0--2K: 262634

Assembly method
Newbler v. May 2013

Additional genomes

Browse all Limulus polyphemus genomes (1)

BioProject

PRJNA20489
An arthropod, commonly known as horseshoe crab, lives in the shores of the Atlantic Coa

Publications

Showing 5 of 9

Mol Biol Evol 2000
The complete mitochondrial DNA sequence of the horseshoe crab Limulus polyphemus
DV Lavrov, et al.

BMC Genomics 2020
The draft genome of horseshoe crab Tachypleus tridentatus reveals its evolutionary scen
Y Zhou, et al.

Genome Biol Evol 2016
Opsin Repertoire and Expression Patterns in Horseshoe Crabs: Evidence from the Genom
Chelicerata)
BA Battelle, et al.

G3 (Bethesda) 2021
Chromosome-level genome assembly of the blue crab, Callinectes sapidus
TR Bachvaroff, et al.

Commun Biol 2021
Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication
W Nong, et al.

View all 9 in PubMed

Annotation details

See full annotation report

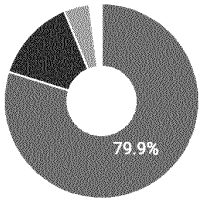
	RefSeq
Provider	NCBI RefSeq
Name	NCBI Annotation Release 101
Date	Aug 28, 2017

	RefSeq
Genes	27,336
Protein-coding	22,873
Non-coding	4,050
Software version	7.4
	View RefSeq annotation

Quality analysis

BUSCO analysis (4.0.2)

Complete 93.7% (S+D)
- Single-copy 79.9%
- Duplicated 13.8%
- Fragmented 4.3%
- Missing 2.0%



C:93.7%[S:79.9%,D:13.8%],F:4.3%,M:2.0%,n:1013
arthropoda_odb10 (1013)

Chromosomes



Download

Chromosome	GenBank	RefSeq	Size (bp)
MT	AF216203.1	NC_003057.1	14,985

Note: This scaffold-level genome assembly includes 286,792 scaffolds and no assembled chromosomes.

Revision history

This record has not been revised

GenBank	RefSeq	Name	Level	Date	Action
GCA_000517525.1	GCF_000517525.1	Limulus_polyphemus-2.1.2	Scaffold	Jan 3, 2014	⋮

FOLLOW NCBI

Connect with NLM

National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894

Web Policies
FOIA
HHS Vulnerability Disclosure

Help
Accessibility
Careers

NLM NIH HHS USA.gov



Nossa *et al. GigaScience* 2014, **3**:9
http://www.gigasciencejournal.com/content/3/1/9

(GIGA)^n SCIENCE

RESEARCH Open Access

Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication

Carlos W Nossa[1,4], Paul Havlak[1], Jia-Xing Yue[1], Jie Lv[1], Kimberly Y Vincent[1], H Jane Brockmann[3] and Nicholas H Putnam[1,2*]

Abstract

Background: Horseshoe crabs are marine arthropods with a fossil record extending back approximately 450 million years. They exhibit remarkable morphological stability over their long evolutionary history, retaining a number of ancestral arthropod traits, and are often cited as examples of "living fossils." As arthropods, they belong to the *Ecdysozoa*, an ancient super-phylum whose sequenced genomes (including insects and nematodes) have thus far shown more divergence from the ancestral pattern of eumetazoan genome organization than cnidarians, deuterostomes and lophotrochozoans. However, much of ecdysozoan diversity remains unrepresented in comparative genomic analyses.

Results: Here we apply a new strategy of combined *de novo* assembly and genetic mapping to examine the chromosome-scale genome organization of the Atlantic horseshoe crab, *Limulus polyphemus*. We constructed a genetic linkage map of this 2.7 Gbp genome by sequencing the nuclear DNA of 34 wild-collected, full-sibling embryos and their parents at a mean redundancy of 1.1x per sample. The map includes 84,307 sequence markers grouped into 1,876 distinct genetic intervals and 5,775 candidate conserved protein coding genes.

Conclusions: Comparison with other metazoan genomes shows that the *L. polyphemus* genome preserves ancestral bilaterian linkage groups, and that a common ancestor of modern horseshoe crabs underwent one or more ancient whole genome duplications 300 million years ago, followed by extensive chromosome fusion. These results provide a counter-example to the often noted correlation between whole genome duplication and evolutionary radiations. The new, low-cost genetic mapping method for obtaining a chromosome-scale view of non-model organism genomes that we demonstrate here does not require laboratory culture, and is potentially applicable to a broad range of other species.

Keywords: Genotyping-by-sequencing (GBS), Genetic linkage mapping, Genome evolution, Limulus polyphemus

Background

Comparative analysis of genome sequences from diverse metazoans has revealed much about their evolution over hundreds of millions of years. The discovery of extensive gene homology across large evolutionary distances has allowed researchers to track chromosome rearrangements and whole genome duplications. The resulting value of whole chromosome sequences presents a challenge for existing whole genome shotgun (WGS) assembly strategies.

Whole genome duplication events were long suspected [1], but only the availability of genome sequences has allowed confirmation of them in fungal, vertebrate, plant and ciliate lineages [2-5]. In contrast, when only a few chordate, insect and nematode genomes were available, conservation of gene linkage (i.e., synteny) and gene order were observed only between closely-related species, and consequently were not expected to be conserved between phyla. As more metazoan genomes have been sequenced, it has become clear that long-range linkage has been conserved over long time scales in many lineages.

Sequencing the genomes of representatives of chordate, mollusk, annelid, cnidarian, placozoan and sponge clades, has identified 17 or 18 ancestral linkage groups (ALGs)

* Correspondence: nputnam@gmail.com
[1]Department of Ecology and Evolutionary Biology, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA
[2]Department of Biochemistry and Cell Biology, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA
Full list of author information is available at the end of the article

BioMed Central

© 2014 Nossa et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

[6-10]. Each of these ALGs consists of a set of ancestral genes whose descendants share conserved synteny in multiple sequenced genomes. These ALGs have been interpreted to correspond to ancestral metazoan chromosomes, and correlations between inferred rates of gene movement between ALGs across the metazoan tree suggest that these ancestral linkage relationships are conserved through the action of selective constraints on a subset of genes [11].

The relatively small number of genomes from anciently distinct metazoan lineages and the fragmented nature of draft genome assemblies still limit both the search for ancient whole genome duplications and the power of the data to constrain models of chromosome-scale genome structure evolution. While WGS sequencing technology and assembly methods are active areas of research and technological development, and have improved at a dramatic pace in recent years, high quality *de novo* assembly of large, complex metazoan genomes remains a difficult and resource-intensive problem. Without genetic or physical maps, or reliance on a high-quality reference genome of a closely-related species, WGS sequencing projects still typically produce assemblies containing thousands of scaffolds, hundreds of scaffolds incorrectly joining sequence from different chromosomes, or both [12].

Next-generation sequencing has greatly reduced the cost of constructing high density genetic maps by eliminating the need to develop and genotype polymorphic markers individually [13]. This has been achieved either by focusing sequence coverage within or adjacent to genomic regions of distinct biochemical character, such as restriction sites with restriction site associated DNA sequencing (RAD-seq) and related methods [14,15], or by combining information across regions using a reference genome sequence [16,17]. While RAD-seq is applicable to organisms lacking a reference genome assembly, it is not directly applicable to comparisons of genome organization across long evolutionary time spans because such comparisons rely on the identification of homologous sequence markers (typically protein-coding genes), which typically have only a small overlap with the restriction-associated markers.

Here we present a genotype-by-sequencing method for constructing a high-density genetic map using low-coverage, low-cost, whole genome sequencing data from the offspring of a wild cross. In this joint assembly and mapping (JAM) approach, the traditionally independent and sequential steps of genome assembly, polymorphic marker identification and genetic map construction are combined. Existing assemblers expect lower densities of sequence polymorphism, deeper coverage, greater computer memory or more aggressive quality trimming that decrease sequence coverage [18-20]. Our current implementation focuses on conservative assembly of short scaffolds sufficient for map construction, but our results suggest that further integration of genetic mapping information within whole genome shotgun assembly methods can be a cost effective way to produce assemblies of large, complex genomes with chromosome-scale contiguity.

We have applied this approach to produce a genetic map of the genome of the Atlantic horseshoe crab, *Limulus polyphemus*. Horseshoe crabs are marine arthropods with a fossil record extending back 450 million years [21]. They exhibit remarkable morphological stability over their long evolutionary history, retaining a number of ancestral arthropod traits [22], and are often cited as examples of "living fossils". *L. polyphemus* has a genome about 90% the size of the human genome. It is an important species from ecological, commercial and conservation perspectives [23], that has been used as a model system for research in behavioral ecology, physiology and development [24]. The map and SNP markers described here will be a resource for the *L. polyphemus* genome project, research in horseshoe crab population biology, and comparisons of metazoan genome organization. By anchoring protein coding genes to this map, we are able to extend analysis of ancestral linkage groups and whole genome duplications to the chelicerate lineage.

Data description

A pair of naturally spawning horseshoe crabs and their eggs were collected from their natural habitat on the beach at Seahorse Key. The larvae were hatched at the lab 4 weeks later from the collecting date. The tissue samples from the third walking legs of the parental horseshoe crabs and 34 larvae were used for DNA extraction, library preparation following manufactures' standard protocols. Two parents and 34 larvae were individually barcoded during library preparation. Illumina paired-end libraries with insert sizes of approximate 300 bp were prepared for each sample. These libraries were pooled together for the subsequent sequncing on the Illumina HiSeq2000 platform at Medical College of Wisconsin Sequencing Service Core Facility. A total of 1.7 billion 100 bp paried-end (PE) reads were obtained after the quality filtering. The total sequencing coverage was estimated as 38.9 × based on the *k*-mer frequency distribution. The raw sequencing data can be retrived from NCBI SRA via NCBI BioProject accession PRJNA187356.

Analyses

Assembly and mapping

The JAM method is designed to produce a combined assembly of polymorphic sequences, tagged by genomic regions with a maximum of one single nucleotide polymorphisms (SNP) per *k*-mer window (Methods). Starting with genomic reads from a mating pair of adult *L. polyphemus* and 34 offspring (100 bp paired-end reads on 300 bp

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

inserts), we analyzed 1.7 billion reads containing at least one high-quality 23-mer.

Fitting Poisson models for unduplicated sequence to the frequencies of filtered 23-mers suggests that 1.1 billion genomic loci are unique at this resolution (Figure 1). This fit assumes that the vast majority of mappable genomic loci have only one or two alleles represented in the parents' four haploid contributions, which gives rise to the four components plotted in Figure 1. Of these sufficiently unique loci, 63% are modeled as homozygous, 27% as paired major-minor alleles, and the remaining 10% as tied allele pairs. The corresponding SNPs, if always at least 23 bases apart, would be 1.6% of bases in the unique loci. Dividing the total number of filtered 23-mers by the modeled homozygous depth of coverage $d = 38.9$ yields an estimated genome size of 2.74 billion bases, consistent with the measured DNA content of 2.8 pg (978 Mb ≃ 1 pg DNA) [25].

We categorize specific 23-mers by their edit distances to others: having no neighbors within a single base substitution (unique tags) or with a single mutually unique one-substitution neighbor ("SNPmer pair" tags). A subset of these, including SNPmer pairs for approximately 7.9 million SNPs, constitute the tags used for contigging and scaffolding. The SNP-mer pairs account for approximately 45% of the modeled fraction of alleles, the others missed from similarity to other sequences (e.g., due to repeats) or distance from each other (because of indels or multiple SNPs per 23-mer).

Chaining these 23-mers together (see Methods) produces an initial 6.6 million contigs, 3.9 million of which are linkable by paired reads for scaffolding. Applying Bambus [26] produces 944,000 scaffolds spanning 1.3 billion bases (Table 1). These scaffolds serve as markers incorporating multiple 23-mer tags, including SNPmer pairs used to identify haplotypes.

After assembly, the mean density of SNPs across the four parental haplotypes in assembled regions was estimated based on read re-alignments to be 7.6 per thousand bases. We jointly inferred the phases of these SNPs and segregation pattern (offspring genotypes) in the mapping cross for each marker in a maximum likelihood framework (Methods). We focused on the 91,320 markers with at least 18 inferred bi-allelic SNPs for constructing the linkage map. These markers grouped into 1,908 high-confidence map bins (i.e., unique segregation patterns, assumed to correspond to loci in the genome uninterrupted by meiotic recombination in the cross [27]. Map bins fell into 32 linkage groups (Figure 2), close to the 26 pairs (2 N = 52) previously found in a cytogenetic analysis of two chromosome spreads [28]. Twenty map bins were removed for having inconsistent positions in the maternal and paternal maps, and 12 were singletons.

To estimate the frequency of incorrect genotype calls as a function of the log likelihood difference between the called and alternative genotype (genotype confidence score), including contributions from uncertainty in SNP-mer identification, assembly and sampling noise, we carried out a simulation of the library pooling and sequencing, *k*-mer assembly and genotype inference protocols, using the sequenced *Ciona intestinalis* genome as a starting point.

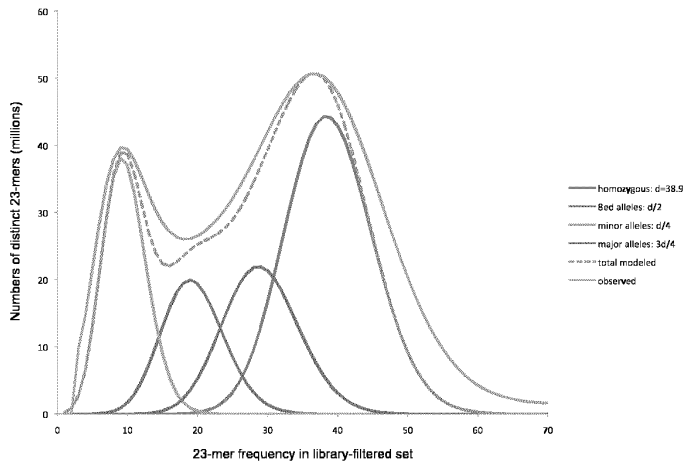


Figure 1 Fitting Poisson distributions to Limulus 23mer frequencies. The distribution of sequenced 23-mers, modeled as sampling genomic loci that are homozygous (single allele shared by all haplotypes), have two alleles that are tied (A and a present in parents as AAxaa or AaxAa), or have two alleles in a major-minor relationship (present in parents as AAxAa or Aaxaa). Alleles whose parental contributions are homozygous, tied, major or minor each have a frequency peak corresponding to their distinctive fraction of the overall depth of genomic sequencing d. Loci sharing simple or repetitive sequence not sufficiently unique at a 23-mer scale contribute to a long tail off the right edge of the plot.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

Table 1 *K*-mer contig and scaffold statistics

Assembled	Count	Total (bp)	Avg. span (bp)	n50 span (bp)
k-mer contigs	6,614,434	1,240,275,515	188	418
Linkable contigs	3,925,844	1,137,576,911	290	460
Initial scaffolds	944,246	1,261,263,172	1336	3047
Reference scaffolds	944,246	1,295,334,515	1372	2930
Reference bases		1,131,458,744	1198	2553

In the simulated *C. intestinalis* data set (Methods), a single stretched exponential distribution provided a good fit to the frequency of genotype calling errors as a function of the call confidence score for scores up to 6, or down to error frequency of approximately 1%. The observed error frequency declined more slowly for higher confidence scores. The minimum x^2 fit used for estimating the genotyping error rate in the *L. polyphemus* map bins was $p_e(s) = a_1\, e^{-\frac{s^{c_1}}{b_1}} + a_2 e^{-\frac{s^{c_2}}{b_2}}$, with parameter values $a_1 = 0.49, b_1 = 2.08, c_1 = 1.26, a_2 = 5.47, b_2 = 0.17, c_2 = 0.16$ (Figure 3).

Applying this model to the *L. polyphemus* marker genome calls, we estimated that the genotype calling error rate in the map bin representative markers was 0.0099. We observed that 51% of adjacent map bin pairs are separated by a single inferred recombination event in the cross, and 94% are separated by three or fewer recombinants in each parent.

Of the 91,320 markers with at least 18 putative SNPs, 84,307 (92%) were assigned to their closest map bins with a threshold of (Methods), for an estimated genome-wide average density of one mapped sequence marker every 32 kb. A mean of 45 markers were mapped to each map bin, and the number of markers mapped was used to estimate the relative physical size of map bins. Approximately 46% of the scaffolds with 12–17 SNPs could be placed with the same threshold, for an additional 32,688 markers, or one marker every 23 kb.

The total length of the scaffolds assigned to map bins was 411 Mb, and they contained 2.67 million bi-allelic SNPs assigned a phase with a posterior probability of at least 0.99. Of these, 72% were inferred to be unique to one of the four parental chromosomes. This is close to the 74% predicted under the finite sites neutral coalescent model given the observed SNP density [29].

Sequence composition and recombination rate

In the scaffolds longer than 1 kb (N = 378,506 and mean length = 2.9 kb), the *G/C* base content was 33.3 ± 2.8%, and the local relative frequency of CpG dinucleotides was bimodally distributed, with about 30% of sequences exhibiting depletion of CpG. TpG and CpA dinucleotides were over-represented on average and their local densities negatively correlated (r = −0.54, p < 2.2e-16) with CpG density, suggesting ongoing germ-line CpG methylation for a fraction of the genome [30].

The mean maternal and paternal recombination rates were estimated to be 1.28 and 0.76 centimorgans per megabase respectively, consistent with expectation based on the negative correlation between recombination intensity and genome size observed in previous studies [31]. We did not observe evidence of segregation distortion for any map bins. The correlation in local recombination rates in two parents across the genome is estimated as $r^2 = 7.1\%$; $p <$ 1e-29, suggesting considerable variation in recombination landscape between two sexes in limulus [32-34]. A positive correlation between local recombination rate and local SNP density was observed ($r^2 = 9.7\%$; $p <$ 1e-40), which is consistent with previous observations in human with comparable correlation coefficient [35].

Ancestral linkage group conservation

We found that 34,942 scaffolds have significant sequence conservation with 10,399 predicted proteins of the tick *Ixodes scapularis*: like *Limulus*, a chelicerate, but one with a well-annotated genome [36]. 6,246 of these hits formed reciprocal best pairs, of which 5,775 (92%) could be placed on the linkage map at a threshold of $p<$. These were used as conserved markers for comparisons of genome organization. When linkage groups were divided into 108 non-overlapping bins of 1,000 markers, 52 had significant ($p < 0.05$, after Bonferroni correction for 1,944 pairwise tests) enrichment in shared orthologs (or "hit") with at least one of eighteen ancestral chordate linkage groups [7]. A hidden Markov model segmentation algorithm [6] identified 40 breakpoints in ALG composition in the linkage groups. Approximately 72% of the genome is spanned by 53 intervening segments that hit one or (for eight of them) two ALGs (Figures 4 and 5). Each of the eighteen ancestral ALGs has at least one hit among the 45 segments with a unique hit to the ALGs.

Whole genome duplications

Whole genome duplication (WGD), or polyploidization is a rare, but dramatic genetic mutation event which doubles the size of a genome and creates a redundant pair of copies from every gene. Because it creates

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

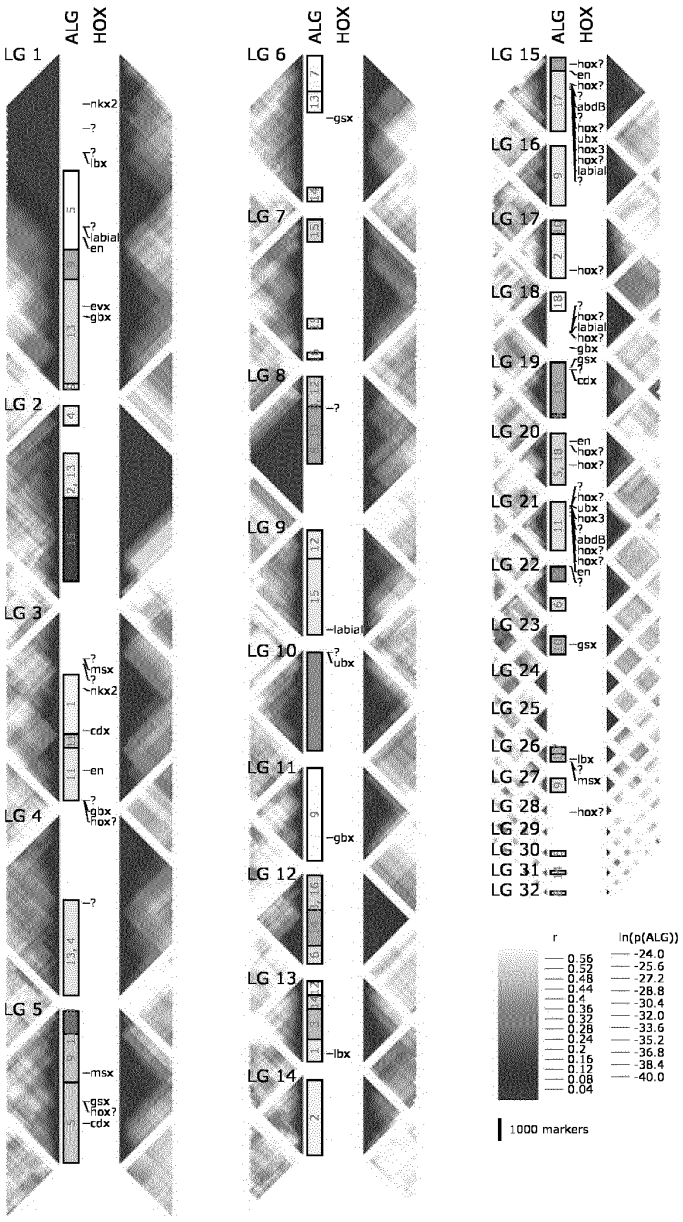


Figure 2 Each of the numbered blocks represents one of the 32 linkage groups of the *L. polyphemus* genetic map, and is composed of four columns: Two bands of the triangular matrices in which the color scale indicates the fraction of samples showing recombination between pairs of markers; maternal recombination frequency is shown on the left, paternal on the right. A column labeled "ALG" indicates segments of significant ($p < 0.05$ in Fisher's Exact Test, after Bonferroni correction for multiple tests) conservation of gene content with ancestral bilaterian linage groups. The column labeled HOX shows the map positions and types of predicted homeobox transcription factor genes. The two color scales are for: recombination frequency between pairs of markers and log p-value for enrichment in gene content with ancestral linkage groups.

redundant copies of genes for entire biochemical pathways and genetic networks, it has been proposed that it creates unique raw material for the evolution of novel biological functions and increased complexity.

Homeobox gene clusters

Homeobox genes encode a large family of transcription factors involved in diverse embryonic patterning and structure formation processes of eukaryotes. As a particular

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

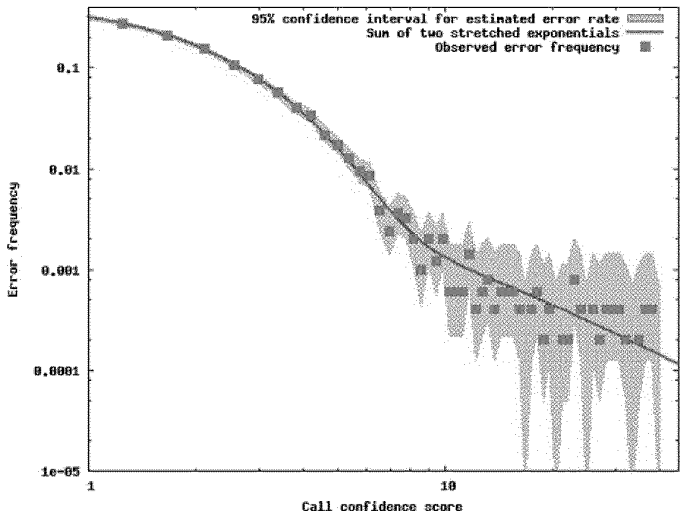


Figure 3 Genotype call error rate as a function of call confidence score for bins of 10,000 calls in simulated *Ciona intestinalis* genome data. The stippled blue region shows 95% confidence intervals of the Bayesian posterior probability distribution of the underlying error rate computed from the Beta distribution $Beta(n_e+1, n_c-n_e+1)$ conjugate the assumed binomial distribution of observed errors, where n_e and n_c are the number of errors and number of calls in each bin respectively. The red curve shows the best fit error model $a_1\, e^{\frac{-c_1}{b_1}} + a_2 e^{\frac{-c_2}{b_2}}$, with parameter values $a_1 = 0.49, b_1 = 2.08, c_1 = 1.26, a_2 = 5.47, b_2 = 0.17, c_2 = 0.16$. χ^2 reduced = 0.82.

subfamily of homeobox genes, the Hox cluster is known to control metazoan body patterning along the anterior-posterior axis. We identified 155 scaffolds with significant homology to predicted chelicerate homeobox gene sequences in public databases. We classified these sequences into homeobox subfamilies (Methods) and placed them on the map by best hit. Two large clusters of Hox genes are found on linkage group (LG) 15 and LG 21, each containing multiple members of the anterior, central and posterior classes. There are also two parahox cluster homologs, each with three homeobox genes: *gsx* and *cdx* orthologs and a third homeobox gene not confidently assigned to a subfamily in our analysis (LG 5 and LG 19). There are two smaller clusters containing multiple hox genes (LG 18 and LG 20), and clusters of other homeobox genes, including members of the *msx*, *lbx*, *nk*, *evx* and *gbx* families (Figure 2).

Genomic distribution of paralogous genes

WGD creates many pairs of duplicate genes or "paralogs". The distinctive features of these genes have been used to

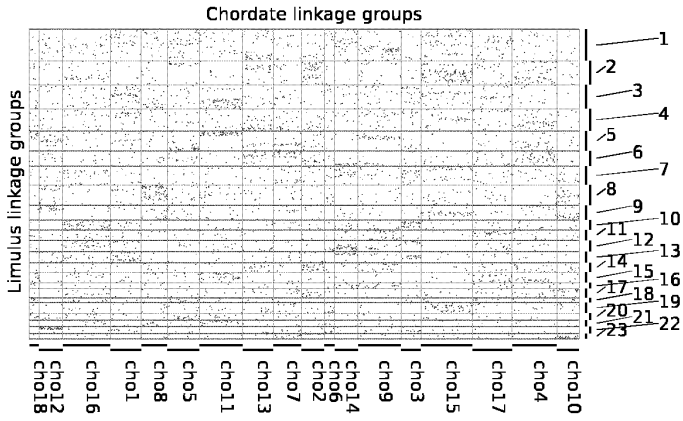


Figure 4 Limulus-Human macro-synteny dot plot. Blue points indicate the position of human genes in reconstructed ancestral chordate ALGs (vertical displacement) and their candidate orthologs in the 30 *L. polyphemus* linkage groups (horizontal displacement).

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

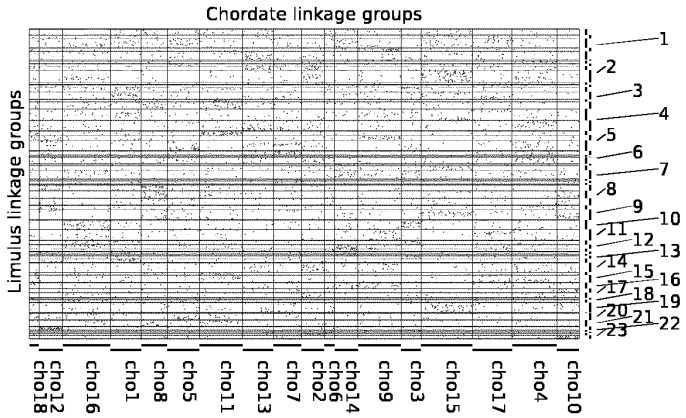


Figure 5 Limulus-Human macro-synteny dot plot as in Figure 4, showing breaks introduced by hidden Markov model segmentation of the linkage groups as vertical lines.

infer WGD events in fungal, vertebrate and plant genomes [2-4]. We examined the genomic distribution of 2,716 pairs of candidate paralogous gene markers in *L. polyphemus* for signatures of WGD. In 45% of these pairs both markers mapped to the same chromosome, compared with 5.3 ± 0.5% in 1,000 datasets with randomly-permuted paralogous gene identities. The mapped positions of pairs within the chromosomes were highly correlated (average $r^2 = 0.81$, and exceeding 0.95 for 8 of the large chromosomes; Figure 6), suggesting that many of the pairs represent recent tandem gene duplicates or single genes fragmented across multiple markers. In the following, these

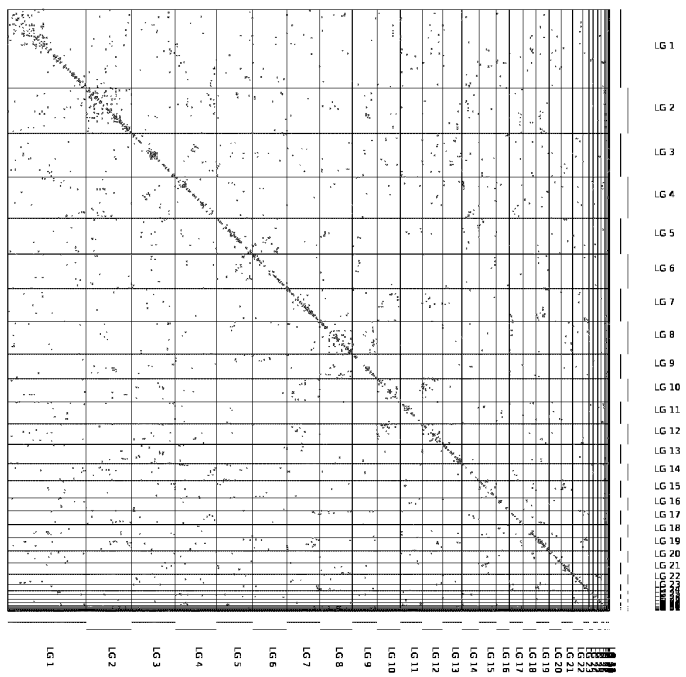


Figure 6 Genomic distribution of candidate paralogs. The map positions of pairs of putatively paralogous protein coding genes within the *L. polyphemus* genome are plotted with blue points. Pairs are biased toward nearby map positions, and therefore concentrated along the diagonal. Also, paralogs split between linkage groups are significantly clustered into "paralogons".

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

same-chromosome paralogs are referred to as "tandem" duplicates.

Inter-chromosomal duplicates are clustered into conserved paralogous micro-synteny blocks (or "paralogons" [3]): there are 25 pairs of loci, each with at least six ($mp = 6$) independent paralog pairs clustered with a maximum gap (max-gap) of 300 markers between adjacent paralogs in each cluster. Paralog pairs are considered independent if they are based on homology to a distinct out-group gene, to guard against relying on either multiple exons of the same gene, or recently-duplicated genes as independent evidence of ancient segmental paralogy. These clusters span 25,044 markers, or 30% of the map, after removing redundancy from paralogons with overlapping footprints. In 1,000 datasets with randomly-permuted paralogous gene identities, the maximum number of such clusters observed was 11; the mean and standard deviation were 3.9 ± 1.0. The observed clustering into paralogons was greater than that in the randomized datasets over a broad range of choices of max-gap and m_p. For example, for max-gap = 100, $mp = 3$ there are 52 clusters vs. 3.5 ± 1.9, range 0–10; for max-gap = 500, $mp = 9$ there are 12 vs. 2.9 ± 1.7, range 0–9. Because of the large proportion of apparent tandem gene duplicates (45%), this randomization scheme increases the number of inter-chromosomal paralog pairs relative to the data, making it a conservative significance test for inter-chromosomal paralog clustering. When genes with tandem duplicates are excluded from the randomizations, the observed number of clusters is greater than the maximum observed in 1000 randomizations for all the combinations of max-gap in the set (100, 200, 300, 400, 500, 600) and *mp* in the set (3, 4, 5, 6, 7, 8, 9). 23 max-gap = 600, $mp = 7$ clusters span 59% of the map, compared to respective mean number and map coverage of 3.3 ± 1.8, and $13 \pm 6\%$ in these randomizations.

Among the marker pairs mapping to different chromosomes, we found a significant excess of pairs relating segments derived from the same ALG relative to randomization controls (247 pairs vs. 102 ± 11, $p < 0.001$ in randomizations of all genes; 202 vs. 46 ± 7 when genes in tandem duplicates are excluded). This pattern is consistent with the creation of these segments by duplication (rather than fission).

The max-gap clusters have a significant amount of overlap among their footprints. For example, the footprints of the max-gap = 600, $mp = 7$ clusters had a total length of 72,072 markers, but a net footprint after redundnacy removal of 49,545 markers. A genomic region *a* which has conserved synteny with two other regions of the genome (*b* and *c*) could arise either from mixing of adjacent regions (through local genome rearragements) with homology to *b* and *c* respectively, or by successive duplications. In the latter case *b* and *c* are also homologous. We examined the relationships among the paralogons for evidence of successive rounds of duplication. We considered a graph in which nodes correspond to merged, non-redun dant paralogon footprint regions. Nodes are connected with edges if a max-gap cluster connects the two nodes. The average clustering coefficient of this graph is equal to the probability that footprints *a* and *c* share a max-gap cluster, given that there are edges (*a*, *b*) and (*b*, *c*) in the graph. We compared the clustering coefficients to those found in random Erdős-Rényi graphs with the same number of nodes and edge probability as the observed graph. We found that the observed data shows significantly more clustering than these random graphs for a wide range of choices of max-gap and m_p. For example, for max-gap 600, m_p =7, the average clustering coefficient is 0.19, while 10,000 random graphs had coefficients of 0.034 ± 0.042, $p = 0.0039$.

Age distribution of paralogous genes

Because WGD events create many paralogs at the same time, they leave characteristic peaks in the age distribution of paralogous genes. In *L. polyphemus* the distribution shows peaks centered at 0.71 and 1.34 substitutions per synonymous site (K_s), values within the approximately linear response range of K_s estimates to WGD age [37] (Figure 7). For comparison, the synonymous site divergence between an Asian horseshoe crab species *Tachypleus tridentatus* and *L. polyphemus* has a mode of 0.35. The common ancestor of these species has been estimated to have lived 114–154 million years ago (MYA), coincident with the opening of the Atlantic ocean [38], suggesting a WGD event 230–310 MYA, and possibly an older one 450–600 MYA.

Discussion

Our results demonstrate that a low cost, combined approach to whole genome sequencing and genetic mapping can be used to efficiently create a very high density genetic recombination map for a non-model organism with a large genome. Because the approach uses genome-wide sequencing, a large number of sequence markers can be anchored to the map, allowing comparisons of genome organization at the chromosome scale over very large evolutionary divergences. The identification of chromosomal segments with significant gene composition homology to each of the chordate ALGs demonstrates that the predominance of fusion and mixing of ancestral linkage groups previously observed in analyzed ecdysozoan genomes [10] is not ancestral to, or universal in the clade.

The map allows quantitative characterization of other features of chromosome-scale organization, such as the correlation between local recombination rate and polymorphism levels. Similar positive correlations between local recombination rate and polymorphism level have been observed in other metazoans including humans

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

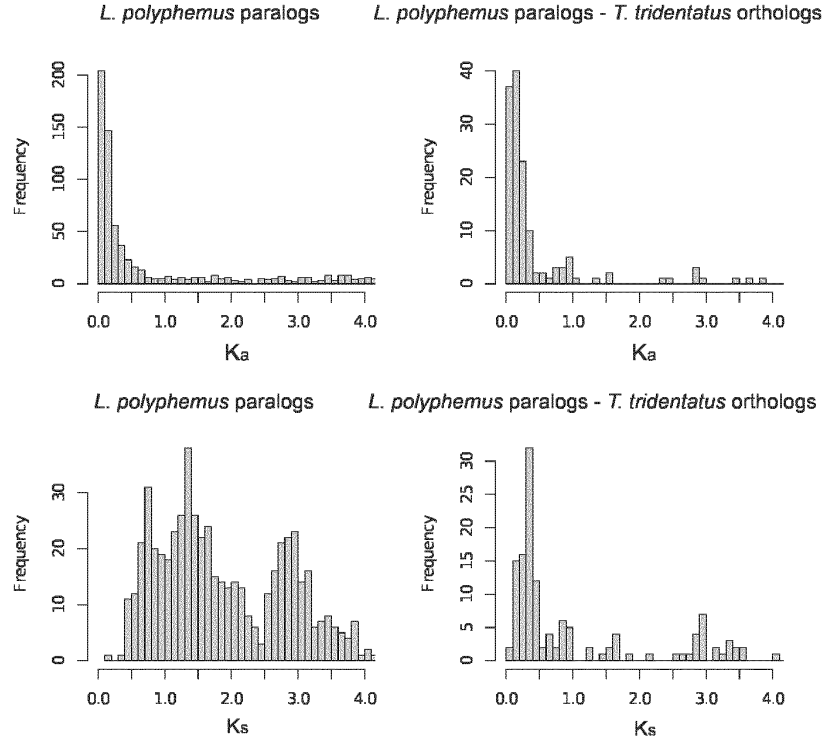


Figure 7 Distribution of estimated non-synonymous (K_a; top) and synonymous (K_s; bottom) and sequence divergence rates for pairs of putative *L. polyphemus* paralogs (left) and *L. polyphemus* - *T. tridentatus* orthologs (right).

[39-41] and plants [42,43]. Future comparisons with more closely related chelicerates will allow tests to distinguish whether these rates are positively correlated with interspecific divergence, consistent with a neutral process of correlated mutation and recombination rates [35]. Alternatively, the association could be explained by hitchhiking and background selection [44].

The enrichment of inter-chromosomal paralog pairs in segments of the same ALG origin is consistent with their creation by duplication (rather than fission), although because small-scale duplication is biased toward local (tandem) duplication, fission of segments could also leave behind an enrichment of paralogs. Such a mechanism, however, would not create the observed organization of paralogs, that is, their clustering into "paralogons". The fact that these paralogons span a large portion of the map (59%) suggests that it was a whole genome duplication, rather than segmental duplications that gave rise to the pattern.

The existence of duplicated hox and parahox clusters on four different chromosomes is highly suggestive of multiple whole genome duplication. Hox clusters have not been found in duplicate copies except in vertebrates where they have been created by whole genome duplication, and have only rarely been subsequently lost.

The double-peaked shape of the distribution of synonymous site divergence between pairs of paralogs, combined with the existence of two small clusters of HOX genes in addition to the two complete HOX clusters suggests that there may have been two rounds of whole genome duplication in the horseshoe crab lineage.

WGDs preceded major species radiations in vertebrates, angiosperms and teleost fish, and the importance of their role in evolution is the subject of long-running debate [1-4]. The discovery of whole genome duplication in an invertebrate, and during horseshoe crabs' long and famously conservative evolutionary history suggests that such events may have been more common than previously assumed in metazoan evolution, and that while they may have provided raw material for adaptive evolution in some cases, they are not evolutionary drivers.

Methods

Joint assembly and mapping (JAM) overview

Barcoded genomic DNA libraries were created, pooled, and sequenced in four lanes on the Illumina HiSeq2000 platform for a mating pair of *L. polyphemus* and 34 offspring.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

The JAM method proceeds through three major phases: 1. The frequencies of DNA sub-sequences of fixed length k (k-mers) are profiled to characterize the quality, uniqueness, polymorphisms and repetition in genomic reads, using software we developed building on work from the Atlas assembler [45]. Allelic pairs of k-mers representing alternate forms of SNPs are identified and tracked through the subsequent steps. 2. Contigs are assembled on a graph of unique k-mers and paired SNP k-mers sampled to reduce memory usage, then ordered and oriented using the Bambus scaffolder [26,46]. Each multi-SNP scaffold is treated as a single marker for the linkage mapping steps 3. The paired SNP k-mers (in each scaffold are combined with the read, mate-pair, and parent- or offspring-library associations of their alleles for haplotype phasing and construction of a high density genetic linkage map. The software is public available as open source software at GitHub [47].

Sampling and sequencing

The parental horseshoe crabs and their eggs were collected from their natural habitat on the beach at Seahorse Key, an island along the west coast of north Florida, on 27 March 2010. This naturally spawning pair were observed as the eggs were being laid and fertilized (fertilization is external in this species, i.e., the eggs are fertilized in the sand under the female as the eggs are being laid). The tissue sample were collected from the third walking legs of this parental pair. We marked where the egg samples were laid and returned a few hours later and dug up the nest, then removed the fertilized eggs. We also have conducted paternity analyses that show that fertilization is by the associated male and not by extraneous sperm that might be at the nesting site (in this case the density of nesting females was low on this day so we know that the eggs we collected were from the pair we observed). Trilobite larvae were reared in plastic dishes as previously described and hatched from the eggs 4 weeks later [48]. Tissue samples and larvae were preserved in RNALater. Genomic DNA purification and library construction were carried out using Qiagen DNAEasy, Illumina TruSeq and Nextera kits, following manufacturers' protocols. Barcoded samples were pooled and sequenced on the Illumina HiSeq2000 platform.

Limulus larvae were processed as follows; each larva, suspended in 100 μL of RNAlater and stored at −80°C in a 1.5 mL Eppendorf tube, was thawed on ice, after which RNAlater was removed. DNA was extracted using the Qiagen DNAEasy kit per manufacturer's protocols. DNA was quantified using picogreen DNA quantitation kit. To prepare TruSeq libraries, DNA was first purified another time using zymo genomic DNA clean columns per manufacturer's protocols. Adult *L. polyphemus* DNA was prepared as above, but using claw tissue rather than whole larvae. All DNA extracts were tested by gel electrophoresis to ensure DNA was not degraded. TruSeq libraries were prepared at University of Georgia's Georgia Genomics Facility. 1–5 μg of sample DNA was subjected to fragmentation using Covaris sonicator. Fragmented DNA was then used for library construction using Illumina TruSeq library prep kits. Libraries were pooled together in equimolar amounts (for 10 larvae) and used for the first sequencing run in separate lanes for the parental and larval pools. For larval samples 11–34, library prep was switched from TruSeq to Nextera kits. Nextera library preparation was performed according to manufacturer's protocol. The Nextera library product was quantified by picogreen, and fragment size distribution was checked by using Lonza flash gel, to ensure that fragment size distribution was between 300–1,000 bp. Sample libraries were pooled in equimolar concentrations and sent for the second sequencing run in two lanes, each on a pool of 12 larvae. Both sequencing runs, comprising four library pools, were performed on the Illumina HiSeq2000 platform at Medical College of Wisconsin Sequencing Service Core Facility.

A total of 1.7 billion Illumina reads qualified for k-mer analysis and assembly by containing at least 23 consecutive q20 bases. The maternal library accounted for 13% of these reads, the paternal library 7.4%, and the 34 offspring libraries for 2.4% on average, 0.64% at minimum.

k-mer decomposition

We determined a lower bound on the k-mer size long enough for a given expectation of uniquenes in a random genome. While increasing k reduced the rate of coincidentally repeated k-mers, it also reduced the effective depth of coverage due to untrimmed errors and edge effects at read ends — and increased the cases of multiple SNPs per k-mer locus, which are not tracked in our current software implementation. We can approximately model a genome-scale string G of random nucleotides as G samples taken with uniform probability from the space of all k-mers (of size $4^k/2$ for odd k-mers treating reverse-complements as same; slightly more for even k). The Poisson distribution then gives the probability that a location in G has its own, unique k-mer (shared with no other location) as

$$e^{-\lambda}, \text{ where } \lambda = \frac{G}{\frac{4^k}{2}} = \frac{2G}{4^k}$$

The probability of a location sharing its *kmer* is then; thus, to limit the maximum rate R of G-locations sharing k-mers, we require $k \geq [\log_4(-2G/\ln(1-R))]$. For example, for a mammalian-scale genome of approximately 3 billion bases, and $R = 0.1\%$, we chose $k \geq 22$. For *Limulus polyphemus*, the Animal Genome Size Database [25] reports an estimated haploid genome size of 2.80 pg and, as each

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

picogram represents almost a billion nucleotide base-pairs of DNA, the mammalian-scale choice of k applies [19,49].

This lower bound ignores chemical and biological sequence biases, so selecting k for a real genome project requires attention to error rates, repeats tandem and interspersed, and genome size, all known vaguely, if at all, before sequencing. Studying the k-mer distributions after sequencing can clarify these genomic properties as we select k to maximize the net yield of candidate k-mer tags, between errors and with at most one SNP location, in sequencing reads. We converted Illumina/Solexa FASTQ format (paying attention to the different quality encodings of the software versions) into FASTA format [50], masking (replacing with 'N') any base with Phred-scale [51] quality below 20, and soft-masking (representing in lower case) other bases with quality below 30. For initial trimming experiments, we varied these quality thresholds as indicated below. We stored k-mers in hash tables with open addressing [52], supporting odd $k <= 31$. We tallied for each k-mer a bit vector for presence or absence in up to 64 sample libraries (36 for *Limulus* parents and offspring), and an overall count of occurrences in all libraries (count limited to $64 - 2k$ bits). Where the k-mer hash would be too large for available memories, we sampled the k-mers using a hash-slicing factor S (must be prime). Representing each k-mer as an integer in, slice s consists of those k-mers whose remainder on division by S is s. We can tabulate one slice for a representative sample of $1/S$ k-mers (for initial estimation of depth of coverage and genome size) or, using S independent jobs, to collect information for k-mers in all slices. Our hash tables stored odd-length k-mers so that reverse-complementary sequences can be combined without the ambiguous orientation of palindromic sequences (e.g., ATCGAT).

After selecting k as described above and making a full tabulation of k-mer counts and bit vectors, we filtered out k-mers not expected to represent genomic sequence. k-mers were required to have three copies in the total sequence set, with at least one copy in the initial run and one in the second run. This was partly to filter out incomplete adapter sequences, which can be difficult to trim, but were different in the two runs.

Extending methods developed for the Atlas assembler [45] to heterozygous sequences, Figure 1 gives a rough decomposition of the k-mer frequency distribution for 23-mers with quality ≥20, minimizing the square of the residuals of k-mer counts on frequencies 3 through 70 while not exceeding the observed counts. Four linked distributions model fractions of the genome as monoallelic or biallelic: homozygous regions with $d = 38.9$-fold coverage (dark blue), minor alleles covered at $d/4$ (green), tied alleles at $d/2$ (red) and major alleles at $3d/4$ (purple). This fit is robust enough to confirm the abundance of major-minor allele pairs (27% of k-loci, vs. 10% for tied alleles), with the broader peaks in the data than in the fitted curves consistent with less uniform sampling (for example, varying coverage of parents and offspring). The Poisson decomposition suggests a density of polymorphisms of 1.2% in major-minor allele pairs, based on dividing the modeled number of such sequenced pairs by k (assuming most polymorphisms are SNPs spaced at least k bases apart), by d (the estimated depth of sequencing) and by the estimated genome size of 2.74 billion bases.

SNPmer identification

The filtered kmer counts, computed in parallel, are loaded into a hash table with additional fields to track kmers that are uniquely within one mismatch of each other. Because this step analyzes all (non-error) k-mers in one table, this requires a single large-memory processor (on the order of 32 GiB).

For each k-mer, we check all its $3k$ one-substitution neighbors. The k-mers are partitioned each into one of three categories: *unique*: having no edit-neighbors within one substitution; *ambiguous*: having either multiple one-substitution neighbors, or one neighbor that has multiple neighbors; or *partnered*: uniquely pairable with exactly one other k-mer differing by one substitution, such k-mers also known from now on as SNPmers or SNPmer pairs. For each SNPmer, we save the position of the substitution, a bitmask for the change (transition, complement, or non-complement transversion), and whether the canonical form of the partner in the table has the same sense or is reverse complemented with respect to this k-mer.

Only partnered and unique k-mers will be further tracked. While this limited method cannot identify k-mers for genomic SNP and non-SNP locations with complete confidence, false pairing or missed pairing should have limited effects, as confirmed by assembly experiments with simulated *Ciona* sequence (see *Error model calibration* in Methods). *False pairing*, due to coincidental similarities or repeats, would combine nodes of the k-mer graph (see below) and cause noise in the scaffolding, haplotype phasing, and linkage analysis. Such misleading links are minimized by the *robust edge* requirements in contigging and scaffolding, described below. *Missed pairing* can happen from indel polymorphisms, SNPs separated by fewer than $k - 1$ positions, failure to sequence minor alleles, or ambiguity due to too many similar k-mers. Ambiguously non-unique k-mers will be skipped over (reducing connectivity of the k-mer graph if there are too many in a row). Where allelic k-mers misidentified as unique cause conflicting edges in the k-mer graph, nodes for unpartnered major alleles will either be chained into contigs with flanking unique sequence or left as orphaned fragments, and unpartnered minor alleles will

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

be left as orphaned fragments. Overall, errors in identifying parnered and unique *k*-mers should shorten contigs and scaffolds and hide linkage, not promote false linkages.

Table 2 shows the totals and percentages of the different kmer categories, counting each SNPmer pair as one *k*-mer. SNPmer pairs account for 16.3% of the putative genomically unique 23mer markers; dividing by 23 gives us the fraction of bases in those markers that are putative SNPs: 0.71% .

Node k-mer selection

To reduce the memory requirements of our *k*-mer assembly graph, we select a approximate one-tenth subset of the SNPmer and unique *k*-mer tags.

In the case of a true SNP at least *k*-1 bases from other SNPs and from gaps in error-free coverage of either allele, there will be *k* covering SNPmer pairs (provided that covering *k*-mers are also uniquely pairable). By taking only SNPmer pairs with the substitution in particular positions, we can reduce the size of the graph and its redundancy. Analyzing the distribution of substituted position for all the SNPmer pairs, we observe an enrichment for substitutions near the ends, probably due to proximity to low-quality sequence. By selecting positions 3, 12 and 21 of 23-base SNPmers, we avoid the most problematic positions and reduce this portion of *k*-mer nodes by a factor of 7.67.

Unlike for SNPmers, there are no canonical positions that identify the unique, unpaired *k*-mers . Several mechanisms have been proposed for sampling *k*-mers in a representative way [53,54]. We use the more pseudo-random hash-slicing rule, already discussed above, to sample a single slice of *k*-mers: those whose integer encodings are congruent to a particular slice number *s*, modulo *S* (the hash slicing factor). We have found that on the finished human genome (results not shown), hash slicing is effectively a Poisson sampling, with sampled *k*-mers spaced according to an exponential distribution.

A caveat in applying hash slicing is that taking the remainder modulo a prime is not very pseudo-random for Mersenne primes (equal to $2^p - 1$ for some p), when *k*-mers are represented in base-4 encoding [52]. We therefore pick a slicing factor of 11, the smallest non-Mersenne prime greater than our SNPmer sampling factor.

Table 2 K-mer categories, counting a SNPmer pair as one *k*-mer

k-mer type	#Distinct	Percentage
No partner/unique	946,431,901	55.48%
Partnered/SNPmer	184,756,149	10.83%
Ambiguous	574,557,296	33.68%
TOTAL	1,705,745,346	

The resulting *k*-mer subset has 86.0 million unique-unpaired *k*-mers and 24.0 million SNPmer pairs, each reduced as predicted, for a total factor of 9.8 reduction in *k*-mer nodes for the next step.

Contigging and scaffolding

Each 23mer tag (unique *k*-mer or SNPmer pair) in the above subset is a node in the *k*-mer graph. Nodes are connected when the corresponding *k*-mers appear consecutively in at least one read of the input (any intervening *k*-mers having been skipped due to sampling or ambiguity). The relative orientation, distance and number of supporting reads of the *k*-mers is stored in the edge. When conflicting distance or relative orientation is observed among different reads for the same pair of *k*-mer nodes, all edges from both nodes in the corresponding direction are ignored in contigging.

The nodes of the *k*-mer graph represent DNA tags and have distinct upstream and downstream ends. One edge at each end of a node is identified as *robust* if supported by a supermajority of the reads for all edges in that direction: the number of supporting reads is greater than or equal to both (1) two plus the sum of the read counts for all other edges in that direction and (2) twice the read count of the next-most supported edge in the same direction. By this construction, a node has at most one robust edge on each end.

A mutually robust edge is defined as one that is robust going in both directions between the two nodes it connects.

Contigs are the connected components of the subgraph consisting of mutually robust edges. Singleton and circular contigs are reported for diagnostic purposes, but ignored in subsequent analysis. Each retained "*k*-mer contig" of Table 1 therefore represents a chain of nodes for SNPmers and unique *k*-mers not shared with other contigs.

After assembly of *k*-mer contigs, we connect them in longer structures using the Bambus scaffolder [26]. Because the contigs do not contain detailed read information, we map templates (read pairs) to contigs based on shared *k*-mer content, and dividing the resulting graph of contigs linked by templates into batches small enough for Bambus to process. Batches are divided so that no contigs in different batches share no templates.

We present templates linking contigs to Bambus using AMOS format [46] for the reads (template ends) mapped to each contig. Reads are included only for the contig with which it shares the most *k*-mers, if the span of those *k*-mers is ≥ *k* and the other read-end of the template similarly qualifies in a different contig. Bambus infers links between contigs by matching template identifiers shared by reads in different "linkable contigs", then produces scaffolds as chains of contigs that are linked with consistent order and orientation.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

Contiguous consensus representations for *k*-mer contigs and scaffolds were generated in two phases. In the first phase, sequence spanned by selected SNPmers and subset *k*-mers (see sections above) are joined together, separated by a number of Ns corresponding to the number of bases not spanned by *k*-mers in the subset. In the second phase, a single pass is made through the read data set, and stretches of Ns that are spanned by single reads are replaced by the sequence of the read.

SNP phase and genotype inference

Each scaffold of the *k*-mer assembly constitutes a candidate marker for mapping. While the depth of sequence coverage on each member of the mapping panel is too low (~1X) to directly infer the genotype of individual members of the mapping pannel at individual SNPs, the tight linkage between SNPs within markers means that learning a sample's genotype at any one reveals it at the others, effectively amplifying the sequence coverage by a factor proportional to the number of SNPs within the marker. This is the same principle exploited in genotype by sequencing (GBS) approaches to genetic mapping in the presence of reference genomes, for example in recombinant inbred lines of reference rice strains [16], and in crosses between *Drosophila* species with sequenced genomes Here we genotype offspring in the context of a cross between two outbred individuals, simultaneously inferring the phases of the SNPs (i.e., which bases appears on each of the four parental chromosomes in the cross). While the data will be insufficient to infer genotypes at many markers, all those where confident inferences can be made can be used to build the linkage map.

For the purposes of genotype inference, a marker is treated as a collection of m SNPs (indexed in the following by $i \in \{1,2,\ldots,m\}$), that have been inferred to be closely linked on the genome via the *k*-mer assembly step. If the four parental chromosomes are labeled *a,b* in one parent and *c,d* in the other, then the genotyping problem is to infer which of the four possible segregation states or genotypes *ac,ad,bc,bd* describes each sample at each marker locus. We index samples with j, and denote a sample genotype by g_j.

We assume that markers are very small compared to a chromosome, and ignore the possibility of a recombination event within individual markers. The data used for inference of the offspring genotypes consist of the number of reads from each barcoded sample j showing each of the four possible DNA bases b at each variable SNP position i, which we denote n_{ij}^b.

If the phase ϕ_i of SNP i were known, *i.e.* which base is present in each of the four parental chromosomes, then a choice of genotype g_j implies a specific homozygous or heterozygous state $s_{ij} \in S = \{AA, CC, TT, GG, AC, AT, AG, CT, CG, TG\}$ for SNP i in sample j. For a given phase and genotype, the likelihood function for a given SNP position in a given sample is given by either a binomial (for homozygous states) or trinomial probability mass function of the read counts, base-calling error rate $\in$, and the site genotype s_{ij}:

$$L(\phi_i, g_j) = p\left(n_{ij}^b \middle| \phi_i, g_j\right) = P_m\left(n_{ij}^b \middle| s_{ij}, \epsilon\right) = \begin{cases} \binom{n}{m} \epsilon^m (1-\epsilon) n-m, & (\text{if } s_{ij} \text{ homozygous}) \\ \frac{n!}{k!l!m!} \left(\frac{2\epsilon}{3}\right)^m \left(\frac{1-\left(\frac{2\epsilon}{3}\right)}{2}\right)^{k+l}, & (\text{if } s_{ij} \text{ heterozygous}) \end{cases}$$

where n is the total number of reads at SNP i; m is the number of observations of bases not in σ_{ij} (i.e., mismatches); k and l are the counts for each of the two bases of σ_{ij} for heterozygous sites.

Likelihood maximization

Searching for an optimal choice of SNP phases ϕ_i and sample genotypes g_j is made difficult by the exponential size of the search space: for segregating bi-allelic SNP sites there are 14 possible phases to consider at each SNP site, so for a mapping panel of only 20 siblings and a marker containing only 10 SNPs, there are combinations to consider. In simulation tests, we found that a variant of expectation maximization (EM), an iterative likelihood maximization method can accurately infer a large proportion of marker genotypes.

To initialize the iteration, the parental samples and a randomly selected offspring are, without loss of generality, assigned genotypes (*a,b*), (*c,d*) and (*a,c*). At each step, we calculate the conditional probability distributions over the possible SNP phases $p(\phi_i)$ given the genotype assignments according to:

$$p^{(t)}(\phi_i) = p\left(\phi_i | g^{(t)}\right) = \prod_{\sigma \in S} p\left(n_{i\sigma}^b | \phi_i, g^{(t)}\right) / \left(\sum_k \prod_{\tau \in S} p\left(n_{k\tau}^b | \phi_k, g^{(t)}\right)\right)$$

where we have labeled the chosen values for the genotypes at iteration t collectively by $g^{(t)}$ $g^{(t)}$ $n_{i\sigma}^b$ is the combined total number of observations of base b at polymorphic SNP i for all samples included at iteration step t which have genotype $s(\phi_i, g_j) = \sigma$.

On each iteration until all samples have been included, a randomly selected sample is added to the set after calculating $p^{(t)}(\phi_i)$. Then the next set of genotype

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

assignments $g^{(t+1)}$ are determined by choosing those that maximize the expected value of the log likelihood:

$$E_{\phi|n,g_j}[\log L(g_j; n, \phi)] = \sum_i p(\phi_i) \log p\left(n_{ij}^b | g_j, \phi_i\right)$$

These steps are repeated until genotypes are being selected for all samples, and the expected log likelihood stops increasing. At the end of the iteration, the likelihood-maximizing genotypes are reported, along with the log likelihood difference between the best and second best choice of genotype for each sample, which provides an indicator of the confidence in genotype call. To gauge convergence, this procedure is repeated 5 times for each marker, with different random choices of initial conditions. Markers which do not identify the same ML genotype multiple times in independent runs are not included among the high confidence genotype calls.

Map bins

Unique marker segregation patterns were included in the set of map bins if they met one of two criteria: (1) at least three independent markers were inferred to have the pattern independently, or (2) the pattern was inferred from at least one marker with at least 20 SNPs such that the mean of the estimated probabilities of the inferred SNP phases was greater than 0.9.

Error model calibration

The sequence of 14 *Ciona intestinalis* autosomes were downloaded from Ensembl [55]. These 14 chromosomes were used as the template in our genome simulation. Based on their sequence length, We used a markovian coalescent simulator macs [56] to generate four haploid samples drawn from a population under neutral Wright-Fisher model with population mutation rate of 0.012 and population recombination rate of 0.0085. Using the *C. intestinalis* genome as the reference sequence, two diploid parental genomes were constructed based on the macs output with realistic SNP and Indel models inferred by several previous studies on the *Ciona* genome [57-59]. We wrote a perl script to simulate the genomes of offspring generated by the cross of the two simulated parents. The software package dwgsim [60] was used to generate Illumina paired-end reads based on our simulated genomes of both parents and offsprings, with the coverage of 20X and 5X respectively.

To estimate the frequency of incorrect genotype calls as a function of the log likelihood difference between the called and alternative genotype, including contributions from uncertainty in SNP-mer identification, assembly, and sampling noise, we carried out a simulation of the *k*-mer assembly and genotype inference protocols Among high-confidence genotype calls, the observed error frequency was a function of call confidence score was well-fit by a sum of two stretched exponential functions, allowing assignment of error probabilities to individual genotype calls.

Linkage group construction

We use the linkage *p*-value p_{ab} between pairs of map bins a and b defined as the minimum over the four possible relabelings r of the maternal and paternal chromosomes of the Binomial *p*-value for the number of matching genotypes:

$$p_{ab} = \min_r \left[1 - \sum_i^{m_r - 1} \binom{n}{i} \frac{1}{2^n}\right]$$

where n is the total number of sample genotype calls (68 in the present case, or 34 in each parent) and m_r is the number of matching genotypes under relabeling r.

We identified map bins with segregation patterns indicating either inconsistent placement in the maternal and paternal maps or genotyping error with a double threshold procedure as follows:

1. Map bins were partitioned into linkage groups by single linkage clustering at a threshold of . $p_{ab} < p_1$.
2. Within each partition, map bins which formed articulation points(i.e., nodes which, if removed, would cause the linkage group to fall apart into two disconnected subgraphs;) in the graph of, $p_{ab} < p_2$, where $p_2 > p_1$.

This procedure identifies map bins which alone account for the merging of what would otherwise be two distinct partitions. We used the following pairs of thresholds p_1, p_2 to identify a total of 20 map bins for exclusion from the map: 10,10; 10,10; 10,10. The remaining markers form locally consistent linkage groups in which all linkages defined at threshold p_1 are corroborated by multiple linkages at p_2, for the above values of p_1 and p_2.

Marker ordering

Markers were ordered within each linkage group using the following protocol. Within each linkage group a consistent labeling of the four parental chromosomes was achieved by constructing a graph G in which nodes correspond to map bins and edges are weighted by linkage *p*-value p_{ab} (as defined above). The local chromosome labels are updated at each map bin as it is reached in a traversal of the minimum spanning tree of G to the labeling r that maximizes p_{ab} along the incident of G used in the traversal. Markers within each linkage group were clustered by hierarchical clustering (marker-marker distance metric: cosine of the angle between the vectors of recombination distances to the other map bins; distance updating method: average linkage) into a binary tree data

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

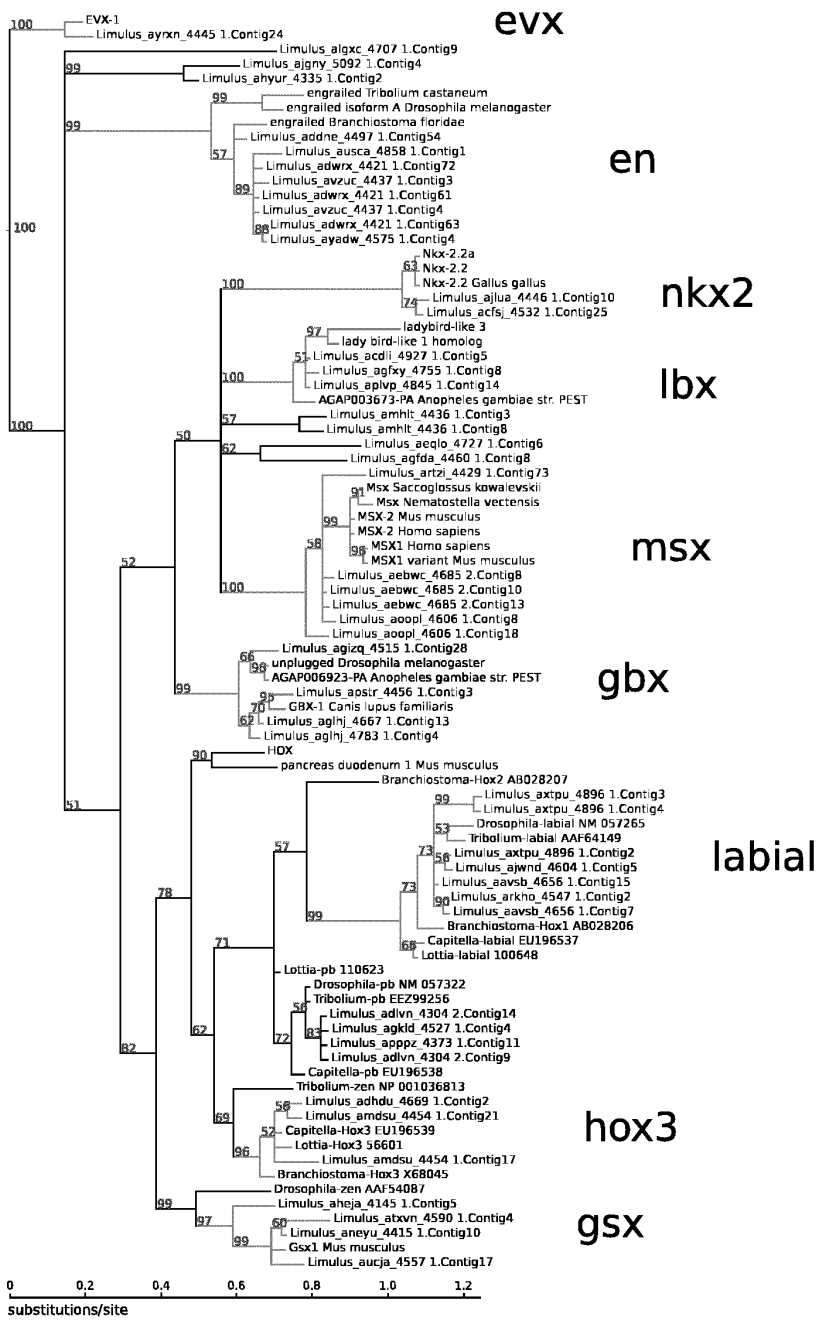


Figure 8 Unrooted phylogenetic tree of homeobox sequences (part 1). Nodes are labeled with Bayesian posterior probabilities. Highly supported partitions used to classify *L. polyphemus* sequences are drawn in red, with the abbreviation for the class shown in large letters. *L. polyphemus* homeobox sequences not grouped into one of these highly supported partitions are assigned to class "? ". For ease of display, a large subtree consisting of HOX and parahox genes has been pruned at the position labeled "HOX", and is shown in Figure 7.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

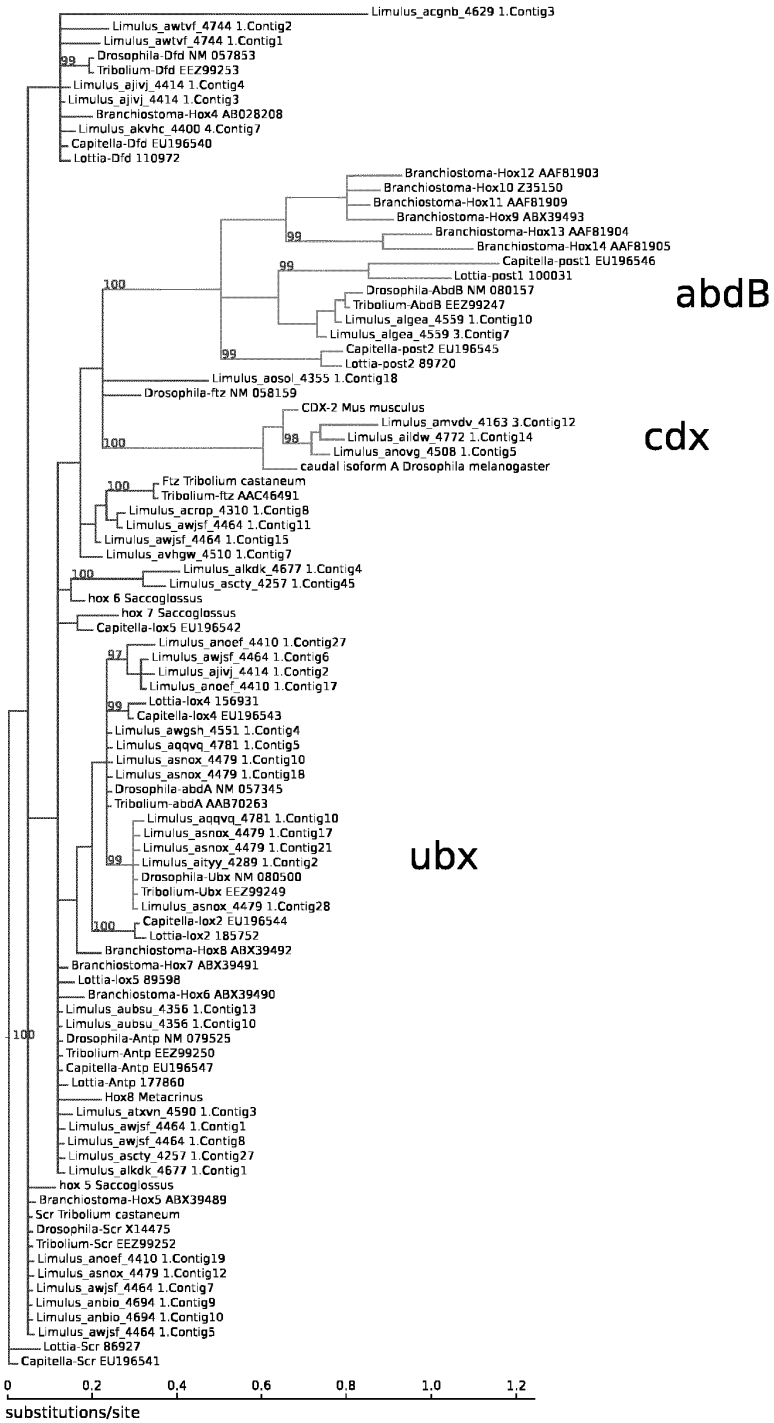


Figure 9 (See legend on next page.)

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

(See figure on previous page.)
Figure 9 Phylogenetic tree of homeobox sequences, part 2. The rooted subtree pruned from the tree in Figure 8. Nodes are labeled with Bayesian posterior probabilities. Highly supported partitions used to classify *L. polyphemus* sequences are drawn in red, with the abbreviation for the class shown in large letters. *L. polyphemus* homeobox sequences not grouped into one of these highly supported partitions are assigned to class "hox?".

structure with leaves representing map bins. A node in the right subtree of the root node was rotated, interchanging its left and right subtrees if its left subtree was not already closer (in average recombination distance) to the markers of the left subtree of the global root; and similarly for nodes in the left subtree of the root. An in-order traversal of the tree generates an ordering of the markers. Finally three reversals of the order of markers in segments of the map were added based on visual inspection of the recombination distance matrix. In the final marker ordering, 51% of adjacent map bin pairs are separated by a single recombination event in the cross, and 94% are separated by three or fewer recombinants in each parent.

Placement of markers on the map

To anchor additional markers to the map, we computed the p_{ab} (see above) between marker a to be placed on the map and each map bin b. Marker a is anchored to the map at the position of the bin b which minimizes p_{ab} if $p_{ab} < 10^{-6}$.

SNP density estimation

Illumina reads were mapped to the assembled scaffold sequences with stampy [61] using default settings. For a sample of 9,228 scaffolds with lengths ranging from 5.0-5.5 kb, sequence variants were called with SAMtools [62] using a variant quality score threshold of 50, and ignoring indel positions.

A SNP density of 0.76% in four haplotypes corresponds to a predicted rate of pairwise sequence differences per site of $\Theta = 0.0042$ under the finite sites model of mutation and the neutral coalescent model of the relationships among sampled alleles [63].

Estimation of local recombination rate

To estimate the local recombination rate for each map bin, we computed the linear regression of map distance in number of markers on physical distance using up to 10 neighboring map bins in each direction along the map (or fewer for bins within 10 map bins of the end of the linkage group). Map distance was calculated from recombination fraction using Haldane's map distance $-\frac{1}{2}\log(1-2r)$ [64].

Ancestral linkage group conservation

To compare the genome organization in *L. polyphemus* to the ancestral metazoan ALGs, we used the reciprocal best blast hit (RBH) orthology criterion in an alignment of the *Ixodes scapularis* predicted proteins [36] to the consensus sequences for the marker scaffolds. *L. polyphemus* scaffolds with RBH of e-value were assigned to the same ancestral bilaterian gene orthology group as their *I. scapularis* ortholog, and thereby with human genes. Regions of the map were tested for enrichment in genes from particular ancestral linkage groups with Fisher's Exact Test, and breakpoints in ancestral linkage group composition were identified using a hidden Markov model, as previously described [6,7].

Homeobox gene modeling

We identified 155 marker scaffolds with a tblastx alignment of e-value to a set of chelicerate homeobox gene sequences downloaded from Genbank using the NCBI online query interface (genbank accessions AF071402.1, AF071403.1, AF071405.1, AF071406.1, AF071407.1, AF085352.1, AF151986.1, AF151987.1, AF151988.1, AF151989.1, AF151990.1, AF151991.1, AF151992.1, AF151993.1, AF151994.1, AF151995.1, AF151996.1, AF151997.1, AF151998.1, AF151999.1, AF152000.1, AF237818.1, AJ005643.1, AJ007431.1, AJ007432.1, AJ007433.1, AJ007434.1, AJ007435.1, AJ007436.1, AJ007437.1, AM419029.1, AM419030.1, AM419031.1, AM419032.1, DQ315728.1, DQ315729.1, DQ315730.1, DQ315731.1, DQ315732.1, DQ315733.1, DQ315734.1, DQ315735.1, DQ315736.1, DQ315737.1, DQ315738.1, DQ315739.1, DQ315740.1, DQ315741.1, DQ315742.1, DQ315743.1, DQ315744.1,

Table 3 Mixture model fits to K_s distribution

N	k	ln(L)	BIC	AIC	Mixture components
1	2	−273.93	559.74	551.87	1.32 ± 0.50
2	5	−259.32	**548.31**	528.64	0.70 ± 0.14 ; 1.45 ± 0.45
3	8	−253.36	554.19	522.71	0.71 ± 0.17 ; 1.34 ± 0.29 ; 2.09 ± 0.19
4	11	−251.41	568.11	524.82	0.74 ± 0.18 ; 1.34 ± 0.20 ; 1.70 ± 0.04 ; 2.02 ± 0.22

N is the number of mixture components, k the number model parameters, ln(L) the log likelihood of the data under the best fit model.
BIC, Bayesian information criterion; AIC,Akaike information criterion.
The BIC score of our selected model is shown in bold.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

EU870887.1, EU870888.1, EU870889.1, HE608680.1, HE608681.1, HE608682.1, HE805493.1, HE805494.1, HE805495.1, HE805496.1, HE805497.1, HE805498.1, HE805499.1, HE805500.1, HE805501.1, HE805502.1, S70005.1, S70006.1, S70008.1, and S70010.1). The reads of each marker (those with best stampy [61] alignment to the scaffold) were reassembled with PHRAP [65], with default parameters. The resulting contigs were aligned to a collection of homeobox-containing protein sequences (genbank accessions NP 001034497.1, NP 001034510.1, AAL71874.1, NP 001034505.1, NP 001036813.1, CAA66399.1, NP 001107762.1, NP 001107807.1, EEZ99256.1, NP 001034519.1, NP 476954.1, NP 032840.1, NP 031699.2, AAI37770.1, EEN68949.1, NP 523700.2, NP 001034511.2, AAK16421.1, and AAK16422.1) with exonerate [66] in protein-to-genome mode. For each contig, the amino acid sequence predicted by the highest-scoring exonerate alignment was used in subsequent phylogenetic analysis, resulting in 104 putative homeobox-containing markers ranging in length from 18 to 147 amino acids.

Phylogenetic analysis of homeobox genes

A multiple sequence alignment of the predicted homeobox sequences combined with a collection of representative sequences from various classes of homeobox genes was constructed with muscle v3.8.31 [67] using default settings. The resulting alignment was trimmed to a 63 amino acid segment spanning the conserved homeodomain, and sequences with more than 50% gaps were removed, leaving 93 predicted *L. polyphemus* homeobox genes in the analysis. Bayesian phylogenetic analysis was carried out on the resulting 178 taxon, 63 amino acid character matrix (See Additional file 1) using MrBayes v3.2.1 [68] using a mixed model of amino acid substituions, gamma-distributed rate variation among sites with fixed shape parameter $\alpha = 1.0$, alignment gaps treated as missing data, 2,000,000 Monte Carlo steps, two independent runs with four Monte Carlo chains, and the initial 25% of sampled trees were discarded as "burn-in". Monte Carlo appeared to reach convergence, with an average standard deviation of the split frequencies of 0.022. The majority-rule consensus of the sampled trees is shown in Figures 8 and 9, and well-supported gene clades (posterior probability greater than 0.95) were used to group the predicted *L. polyphemus* genes into classes. The table in Additional file 1 lists the reassembled marker contigs, their inferred hox gene class, and maximum likelihood map positions. Predicted genes were anchored to the map as described above.

Genomic distribution of paralogs

We identified 2,716 pairs of *Limulus* markers that can both be placed on the map and have their best translated alignment to the same *Ixodes scapularis* gene. (*I. scapularis* genes with more than five best-hit markers were excluded from seeding such pairs.) To estimate the synonymous sequence divergence between pairs of candidate *L. polyphemus* paralogous gene pairs and *L. polyphemus* genes and their *T. tridentatus* orthologs, we constructed codon alignments of predicted coding sequence for estimation of synonymous sequence divergence. Conserved clusters of paralogs were identified using a variant of the "max-gap"

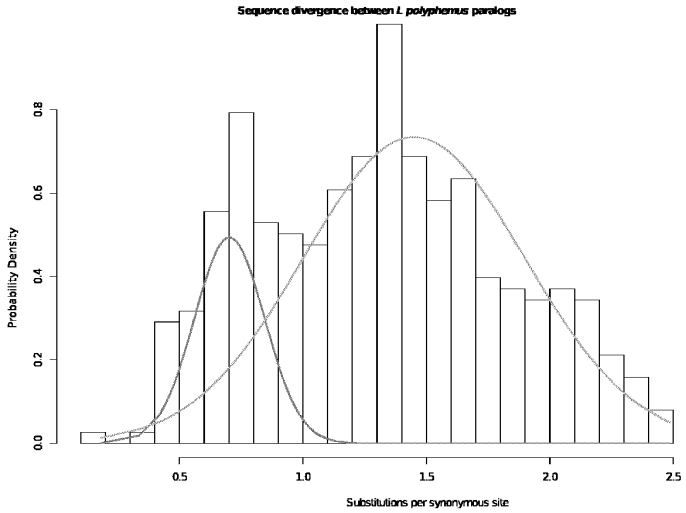


Figure 10 Two component mixture model fit to the K_s peak on the range $0 <= K_s <= 2.5$. The best-fitting model was selected by the Bayesian Information Criterion (Table 3). The component means are 0.7 and 1.45 substitutions per site. The position of the peak at lowest K_s was not sensitive to the addition of more mixture components.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

criterion [3] in which two genes are placed in the same cluster if they and their paralogs lie within threshold distance.

K_a and K_s estimation for paralogs and T. tridentatus orthologs

Figure 6 shows the distribution across the map of pairs of candidate paralogs. To estimate the synonymous sequence divergence between pairs of candidate *L. polyphemus* paralogous gene pairs, and between *L. polyphemus* genes and their *T. tridentatus* orthologs, we followed the following protocol.

1. Reassemble reads from each marker with PHRAP [65] , and create a predicted coding sequence using exonerate, as described for the annotation of homeobox gene models (see above).
2. Combine the exonerate alignments of codons to amino acids to create an alignment of codons for either a pair of *L. polyphemus* sequences, or for a *L. polyphemus* - *T. tridentatus* sequence pair.
3. Use the method of Yang and Nielsen [69] to estimate the synonymous and non-synonymous substitution rates K_a and K_s, as implemented in the KaKsCalculator package [70].
4. Discard estimates based on fewer than 30 sites (30 synonymous sites for estimates of K_s, non-synonymous sites for K_a).

GenBank accessions for *Tachypleus tridentatus* mRNA clones: JQ966943, AB353281, AB353280, HM156111, HQ221882, HQ221883, HQ221881, HQ386702, HM852953, TATTPP, TATPROCLOT, FN582225, FN582226, AF467804, AF227150, GQ260127, AF264067, AF264068, AB353279, AB005542, TATLICI, TATTGL, TATCFGB, TATLFC1, TATLFC2, AB201713, TATCFGA, TATLICI2, CS423581, CS423579, AB028144, AB201778, AB201776, AB201774, AB201772, AB201770, AB201768, AB201766, AB201779, AB201777, AB201775, AB201773, AB201771, AB201769, AB201767, AB201765, AB105059, AB002814, AX763473, TATCFBP, AB076186, AB076185, X04192, TATHCLL, AB037394, AB019116, AB019114, AB019112, AB019110, AB019108, AB019106, AB019104, AB019102, AB019100, AB019098, AB019096, AB019117, AB019115, AB019113, AB019111, AB019109, AB019107, AB019105, AB019103, AB019101, AB019099, AB019097, AB023783, AB024738, AB024739, AB024737, AB017484, D87214, D85756, D85341.

Figure 7 shows the distribution of K_a and K_s for paralogs and *T. tridentatus* orthologs. To estimate the number and age of peaks in the un-saturated range [37] of the K_s distribution (and of putative WGD events), we fit a series of univariate normal mixture models, with 1, 2, 3, and 4 components to the paralog K_s distribution in the range $0 <= K_s <= 2.5$ and selected the best model on the basis of Bayesian Information Criterion (BIC) (Table 3). The best model had two components, with means at 0.7 and 1.45 substitutions per site. The position of the peak at lowest K_s was not sensitive to the addition of more mixture components. Figure 10 shows comparison of the distribution and the components of the best-fitting model. Gaussian mixture models were estimated in R with mixtools [71].

Availability of supporting data

The raw sequencing reads are currently being submitted through the NCBI SRA and are accessible via NCBI BioProject accession PRJNA187356.

The data sets supporting the results of this article are available in the *GigaScience* GigaDB repository [72].

Additional file

Additional file 1: The amino acid character matrix used for the phylogenetic analysis of homeobox genes.

Abbreviations

ALG: Ancestral linkage groups; AIC: Akaike information criterion; BIC: Bayesian information criterion; bp: Basepair; EM: Expectation maximization; GBS: Genotyping-by-sequencing; JAM: Joint assembly and mapping; LG: Linkage group; Max-gap: Maximum gap; Mb: Megabase; MYA: Million years ago; PE: Paired-end; RAD-Seq: Restriction site associated DNA sequencing; RBH: Reciprocal best blast hit; SNP: Single nucleotide polymorphism; WGD: Whole-genome duplication.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NHP conceived and led the project. All authors wrote the paper. PH, NHP and J-XY wrote software. NHP, PH, J-XY and CWN carried out sequence analysis. JB collected and raised samples. CWN extracted genomic DNA and created the libraries for sequencing. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by the National Science Foundation (EF-0850294 and IOB 06–41750), the Beckman Young Investigator Program, the University of Florida Division of Sponsored Research, the Department of Biology, and the UF Marine Laboratory at Seahorse Key. We thank the three reviewers, Dr. Hugues Roest Crollius, Dr. Stephen Richards and Dr. Brian Eads, for their valuable comments which greatly helped to improve the quality of this paper.

Author details

[1]Department of Ecology and Evolutionary Biology, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA. [2]Department of Biochemistry and Cell Biology, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA. [3]Department of Biology, University of Florida, P.O. Box 11-8525 Gainesville, FL 32611-8525, USA. [4]Current address: Gene by Gene, Ltd, Houston, TX 77008, USA.

Received: 23 October 2013 Accepted: 23 April 2014
Published: 14 May 2014

References

1. Ohno S: *Evolution by Gene Duplication*. Springer-Verlag; 1970.
2. Wolfe KH, Shields DC: **Molecular evidence for an ancient duplication of the entire yeast genome.** *Nature* 1997, **387**:708–713.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

3. McLysaght A, Hokamp K, Wolfe KH: **Extensive genomic duplication during early chordate evolution.** *Nat Genet* 2002, **31**:200–204.
4. Simillion C, Vandepoele K, Montagu MCEV, Zabeau M, de Peer YV: **The hidden duplication past of Arabidopsis thaliana.** *Proc Natl Acad Sci* 2002, **99**:13627–13632.
5. Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, Arnaiz O, Billaut A, Beisson J, Blanc I, Bouhouche K, Câmara F, Duharcourt S, Guigo R, Gogendeau D, Katinka M, Keller A-M, Kissmehl R, Klotz C, Koll F, Mouël AL, Lepère G, Malinsky S, Nowacki M, Nowak JK, Plattner H, *et al*: **Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia.** *Nature* 2006, **444**:171–178.
6. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS: **Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization.** *Science* 2007, **317**:86–94.
7. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, Benito-Gutierrez E, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, *et al*: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453**:1064–1071.
8. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS: **The Trichoplax genome and the nature of placozoans.** *Nature* 2008, **454**:955–960.
9. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, *et al*: **The Amphimedon queenslandica genome and the evolution of animal complexity.** *Nature* 2010, **466**:720–726.
10. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, de Jong P, Grimwood J, Chapman JA, Shapiro H, Aerts A, Otillar RP, Terry AY, Boore JL, Grigoriev IV, Lindberg DR, Seaver EC, Weisblat DA, Putnam NH, Rokhsar DS: **Insights into bilaterian evolution from three spiralian genomes.** *Nature* 2013, **493**:526–531.
11. Lv J, Havlak P, Putnam N: **Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes.** *BMC Bioinformatics* 2011, **12**(Suppl 9):S11.
12. Earl D, Bradnam K, John JS, Darling A, Lin D, Fass J, Yu HOK, Buffalo V, Zerbino DR, Diekhans M, Nguyen N, Ariyaratne PN, Sung W-K, Ning Z, Haimel M, Simpson JT, Fonseca NA, Birol İ, Docking TR, Ho IY, Rokhsar DS, Chikhi R, Lavenier D, Chapuis G, Naquin D, Maillet N, Schatz MC, Kelley DR, Phillippy AM, Koren S, *et al*: **Assemblathon 1: A competitive assessment of de novo short read assembly methods.** *Genome Res* 2011, **21**:2224–2241.
13. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML: **Genome-wide genetic marker discovery and genotyping using next-generation sequencing.** *Nat Rev Genet* 2011, **12**:499–510.
14. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES: **An SNP map of the human genome generated by reduced representation shotgun sequencing.** *Nature* 2000, **407**:513–516.
15. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA: **Rapid SNP discovery and genetic mapping using sequenced RAD markers.** *PLoS ONE* 2008, **3**:e3376.
16. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B: **High-throughput genotyping by whole-genome resequencing.** *Genome Res* 2009, **19**:1068–1076.
17. Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL: **Multiplexed shotgun genotyping for rapid and efficient genetic mapping.** *Genome Res* 2011, **21**:610–617.
18. Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS: **Meraculous: De novo genome assembly with short paired-end reads.** *PLoS ONE* 2011, **6**:e23501.
19. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB: **ALLPATHS: De novo assembly of whole-genome shotgun microreads.** *Genome Res* 2008, **18**:810–820.
20. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB: **High-quality draft assemblies of mammalian genomes from massively parallel sequence data.** *Proc Natl Acad Sci* 2011, **108**:1513–1518.
21. Rudkin DM, Young GA: **Horseshoe crabs – an ancient ancestry revealed.** In *Biol Conserv Horseshoe Crabs.* Edited by Tanacredi JT, Botton ML, Smith D. New York: Springer; 2009:25–44.
22. Fisher DC: **The Xiphosurida: archetypes of Bradytely?** In *Living Foss.* Edited by Eldrege N, Stanley SM. New York: Springer; 1984:196–213.
23. Berkson J, Shuster CN Jr: **The horseshoe crab: the battle for a true multiple-use resource.** *Fisheries* 1999, **24**:6–10.
24. Shuster CN Jr, Barlow RB, Brockmann HJ (Eds): *The American Horseshoe Crab.* Cambridge, MA: Harvard University Press; 2003.
25. Gregory TR: *Animal Genome Size Database.* http://www.genomesize.com.
26. Pop M, Kosack DS, Salzberg SL: **Hierarchical scaffolding with bambus.** *Genome Res* 2004, **14**:149–159.
27. Van Os H, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, de Jong W, van Koert P, Lefebvre V, Milbourne D, Ritter E, van der Voort JNAMR, Rousselle-Bourgeois F, van Vliet J, Waugh R, Visser RGF, Bakker J, van Eck HJ: **Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map.** *Genetics* 2006, **173**:1075–1087.
28. Sekiguchi K: *Biology of Horseshoe Crabs.* Tokoyo: Science House Co., Ltd.; 1988:50–68.
29. Cartwright RA, Hussin J, Keebler JEM, Stone EA, Awadalla P: **A family-based probabilistic method for capturing de novo mutations from high-throughput short-read sequencing data.** *Stat Appl Genet Mol Biol* 2012, 11(2).
30. Bird AP: **DNA methylation and the frequency of CpG in animal DNA.** *Nucleic Acids Res* 1980, **8**:1499–1504.
31. Lynch M: **The origins of eukaryotic gene structure.** *Mol Biol Evol* 2006, **23**:450–468.
32. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: **Comprehensive human genetic maps: individual and sex-specific variation in recombination.** *Am J Hum Genet* 1998, **63**:861–869.
33. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K: **A high-resolution recombination map of the human genome.** *Nat Genet* 2002, **31**:241–247.
34. Coop G, Przeworski M: **An evolutionary view of human recombination.** *Nat Rev Genet* 2007, **8**:23–34.
35. Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M: **A neutral explanation for the correlation of diversity with recombination rates in humans.** *Am J Hum Genet* 2003, **72**:1527–1535.
36. VectorBase: **Ixodes scapularis annotation, IscaW1.** https://www.vectorbase.org/organisms/ixodes-scapularis.
37. Vanneste K, Van de Peer Y, Maere S: **Inference of genome duplications from age distributions revisited.** *Mol Biol Evol* 2013, **30**:177–190.
38. Obst M, Faurby S, Bussarawit S, Funch P: **Molecular phylogeny of extant horseshoe crabs (Xiphosura, Limulidae) indicates Paleogene diversification of Asian species.** *Mol Phylogenet Evol* 2012, **62**:21–26.
39. Begun DJ, Aquadro CF: **Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster.** *Nature* 1992, **356**:519–520.
40. Cutter AD, Payseur BA: **Selection at linked sites in the partial selfer caenorhabditis elegans.** *Mol Biol Evol* 2003, **20**:665–673.
41. Nachman MW: **Single nucleotide polymorphisms and recombination rate in humans.** *Trends Genet* 2001, **17**:481–485.
42. Stephan W, Langley CH: **DNA polymorphism in lycopersicon and crossing-over per physical length.** *Genetics* 1998, **150**:1585–1593.
43. Roselius K, Stephan W, Städler T: **The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species.** *Genetics* 2005, **171**:753–763.
44. Andolfatto P, Przeworski M: **Regions of lower crossing over harbor more rare variants in African populations of drosophila melanogaster.** *Genetics* 2001, **158**:657–665.
45. Havlak P, Chen R, Durbin KJ, Egan A, Ren Y, Song X-Z, Weinstock GM, Gibbs RA: **The atlas genome assembly system.** *Genome Res* 2004, **14**:721–732.
46. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M: **Next generation sequence assembly with AMOS.** *Current Protocols in Bioinformatics* 2011, **33**:11.8.1–11.8.18.
47. **The JAM-pipeline.** https://github.com/putnamlab/jam-pipeline.
48. Johnson SL, Brockmann HJ: **Costs of multiple mates: an experimental study in horseshoe crabs.** *Anim Behav* 2010, **80**:773–782.

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

49. Mullikin JC, Ning Z: **The phusion assembler.** *Genome Res* 2003, **13**:81–90.
50. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM: **The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants.** *Nucleic Acids Res* 2010, **38**:1767–1771.
51. Ewing B, Green P: **Base-calling of automated sequencer traces usingphred. ii. error probabilities.** *Genome Res* 1998, **8**:186–194.
52. Knuth DE: *Searching and Sorting*, Volume 3. Reading, MA: Addison-Wesley; 1973 [*The Art of Computer Programming*].
53. Roberts M, Hayes W, Hunt BR, Mount SM, Yorke JA: **Reducing storage requirements for biological sequence comparison.** *Bioinformatics* 2004, **20**:3363–3369.
54. Ye C, Ma ZS, Cannon CH, Pop M, Yu DW: **Exploiting sparseness in de novo genome assembly.** *BMC Bioinformatics* 2012, **13**(6):S1.
55. **Ensembl.** http://www.ensembl.org/index.html.
56. Chen GK, Marjoram P, Wall JD: **Fast and flexible simulation of DNA sequence data.** *Genome Res* 2009, **19**:136–142.
57. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, Tomaso AD, Davidson B, Gregorio AD, Gelpke M, Goodstein DM, Harafuji N, Hastings KEM, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang H-G, Awazu S, Azumi K, *et al*: **The draft genome of ciona intestinalis: insights into chordate and vertebrate origins.** *Science* 2002, **298**:2157–2167.
58. Haubold B, Pfaffelhuber P, Lynch M: **mlRho – a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes.** *Mol Ecol* 2010, **19**:277–284.
59. Small KS, Brudno M, Hill MM, Sidow A: **Extreme genomic variation in a natural population.** *Proc Natl Acad Sci* 2007, **104**:5698–5703.
60. **dwgsim.** https://github.com/nh13/DWGSIM/releases.
61. Lunter G, Goodson M: **Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads.** *Genome Res* 2011, **21**:936–939.
62. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**:2078–2079.
63. Yang Z: **Statistical properties of a DNA sample under the finite-sites model.** *Genetics* 1996, **144**:1941–1950.
64. Haldane JBS: **The combination of linkage values and the calculation of distances between the loci of linked factors.** *J Genet* 1919, **8**:299–309.
65. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8**:195–202.
66. Slater GS, Birney E: **Automated generation of heuristics for biological sequence comparison.** *BMC Bioinformatics* 2005, **6**:31.
67. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792–1797.
68. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572–1574.
69. Yang Z, Nielsen R: **Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models.** *Mol Biol Evol* 2000, **17**:32–43.
70. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J: **KaKs calculator: calculating Ka and Ks through model selection and model averaging.** *Genomics Proteomics Bioinformatics* 2006, **4**:259–263.
71. Benaglia T, Chauveau D, Hunter, David R, Young, Derek S: **mixtools: An R package for analyzing finite mixture models.** *J Stat Softw* 2009, **32**:1–29.
72. Nossa CW, Havlak P, Yue JX, Lv J, Vincent KY, Brockmann HJ, Putnam NH: **Supporting materials from "Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication".** 2014. GigaScience Database. http://dx.doi.org/10.5524/100091.

doi:10.1186/2047-217X-3-9
Cite this article as: Nossa *et al.*: Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *GigaScience* 2014 **3**:9.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Downloaded from https://academic.oup.com/gigascience/article/3/1/2047-217X-3-9/2682910 by guest on 27 June 2023

Seikagaku Corporation Exhibit 2004 - Page 205 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

An official website of the United States government
Here's how you know

Log in

Nucleotide

GenBank

PREDICTED: Limulus polyphemus limulus clotting factor C-like (LOC106477041), mRNA

NCBI Reference Sequence: XM_013937646.1

This sequence has been updated. See current version.

FASTA Graphics

Go to:

```
LOCUS       XM_013937646            1591 bp    mRNA    linear   INV 09-SEP-2015
DEFINITION  PREDICTED: Limulus polyphemus limulus clotting factor C-like
            (LOC106477041), mRNA.
ACCESSION   XM_013937646
VERSION     XM_013937646.1
DBLINK      BioProject: PRJNA238073
KEYWORDS    RefSeq.
SOURCE      Limulus polyphemus (Atlantic horseshoe crab)
  ORGANISM  Limulus polyphemus
            Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Chelicerata;
            Merostomata; Xiphosura; Limulidae; Limulus.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_013678219.1) annotated using gene prediction method: Gnomon.
            Also see:
                Documentation of NCBI's Annotation Process

            [WARNING] On Aug 31, 2017 this sequence was replaced by
            XM_013937646.2.

            ##Genome-Annotation-Data-START##
            Annotation Provider         :: NCBI
            Annotation Status           :: Full annotation
            Annotation Version          :: Limulus polyphemus Annotation
                                           Release 100
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 6.4
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..1591
                     /organism="Limulus polyphemus"
                     /mol_type="mRNA"
                     /db_xref="taxon:6850"
                     /chromosome="Unknown"
                     /sex="male"
                     /tissue_type="muscle"
                     /country="USA: Woods Hole, MA"
                     /collection_date="Jan-2008"
     gene            1..1591
                     /gene="LOC106477041"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Gnomon. Supporting evidence
                     includes similarity to: 99 Proteins, and 100% coverage of
                     the annotated genomic feature by RNAseq alignments,
                     including 5 samples with support for all annotated
                     introns"
                     /db_xref="GeneID:106477041"
     CDS             391..1086
                     /gene="LOC106477041"
                     /codon_start=1
```


Seikagaku Corporation Exhibit 2004 - Page 206 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

```
                     /product="limulus clotting factor C-like"
                     /protein_id="XP_013793100.1"
                     /db_xref="GeneID:106477041"
                     /translation="MWFLQCGGALLNEKWIITAAHCVTYSATAEIIDPSQFKFYLGKY
                     YRDDSKDDDYVQVREAIEIHVNPNYDPGNLNFDIALIQLKTSVALTTRVQPICLPTDL
                     TTRENLKEGALAVVTGWGLNENNTYSEMIQQAVLPVVAASTCEQGYQDSGLPLTVTEN
                     MFCAGYKQGRYDACSGDSGGPLVFADDSRTDRRWVLEGIVSWGSPNGCGKSNQYGGFT
                     KVNVFLSWIRQFI"
ORIGIN
        1 aaaaagttaa taccagcctc ttttttttcaa taatctaggt tgccacggtt acccctccct
       61 ttagccagag cagccggacc tcctccaaaa cctagatccg caccaccctc tactgtggac
      121 ctgtcttcca aggtcaaact gcctgaaggt cattaccggg tggggtctca agccatttac
      181 acgtgcgagt caagatacta cgaactgctt ggatctcaag gtagaagatg cgactctaat
      241 ggaaagtgga gtggtcgacc agcaagctgt ataccagttt gtggacggtc agactctccc
      301 cgttctcctt tcatcgtcaa tggaaattcc accgaaatag gtcagtggcc gtggcaggca
      361 ggaatctcca gatggcttgc agatcataat atgtggtttc ttcagtgtgg aggagcccta
      421 ctgaatgaga aatggatcat tactgcagcc cactgtgtca cctactctgc tactgccgag
      481 atcattgacc caagtcagtt taaattctac ctgggcaaat actatcgaga tgacagtaag
      541 gatgatgact acgtacaagt aagagaggct atcgagatcc atgtgaatcc taactacgat
      601 cctggaaatc tcaactttga catagccctg attcaactga agacttctgt tgctctgacc
      661 acacgagtgc aaccaatatg tctgcctact gatctcacta caagagaaaa cctgaaagag
      721 ggagcgttag cggtggtgac aggatggggt ttgaatgaaa acaacacata ttcagagatg
      781 attcagcaag ccgttctgcc tgttgttgca gcaagcacct gtgaacaagg ataccaggac
      841 tcgggcttgc cactgacagt gacagagaac atgttctgtg caggttacaa gcaggggcgc
      901 tatgatgcct gcagtggaga cagtggagga ccattagtgt ttgctgatga ttcccgcacc
      961 gataggcggt gggtcctgga agggatcgtc agctggggca gccccaatgg atgtggcaag
     1021 tctaaccagt atgggggctt cactaaagtt aacgtttttc tatcgtggat tcggcagttc
     1081 atttgaaaac gttacaaatg ctttaatctt gctttaattg tactcataag aagggtaatc
     1141 aagtaggtca tcaatgtcat tatgtgttat ttgtttttac aaatggctat tttagtcagt
     1201 gaatgagaat attcatctga aaattttacc ttcaattttt tattctaccc ttttctatga
     1261 aaatgtaagt gttttgtata tcttctatac atatatgaaa attctgaatt tttgcataaa
     1321 tttgatttgg tttggcttat atgttaatat gtacaataca tattgctcct tatttttata
     1381 ttataaatag tttatcctca gttctgcctg tgatgactaa ataaaacact gaacaacgat
     1441 tcaaattcta gtagagcatg aaatgtatta tcagttacct taagtaaaag agttagcttc
     1501 acttatcaaa ataaaatgtc tagtaacttc tcagactctc ttaccctgac tggtcgttca
     1561 ttcattcatc acatgataaa aaagtaataa t
//
```


Seikagaku Corporation Exhibit 2004 - Page 207 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

NCBI Limulus polyphemus Annotation Release 100

The RefSeq genome records for Limulus polyphemus were annotated by the NCBI Eukaryotic Genome Annotation Pipeline, an automated pipeline that annotates genes, transcripts and proteins on draft and finished genome assemblies. This report presents statistics on the annotation products, the input data used in the pipeline and intermediate alignment results.

The annotation products are available in the sequence databases and on the FTP site.

This report provides:

- Annotation Release information: The name of the release, important dates, the software version
- Assemblies: A brief description of the annotated assembly(ies)
- Gene and feature statistics: The counts and characteristics of the annotated features
- Alignment of the annotated proteins to a set of high-quality proteins: The number of annotated proteins with hits to a set of high-quality proteins
- Masking of genomic sequence: How much of the genome was masked
- Transcript and protein alignments: The number and type of evidence retrieved from public databases and used for gene prediction

For more information on the annotation process, please visit the NCBI Eukaryotic Genome Annotation Pipeline page.

Annotation Release information

This annotation should be referred to as NCBI Limulus polyphemus Annotation Release 100

Annotation release ID: 100
Date of Entrez queries for transcripts and proteins: Sep 2 2015
Date of submission of annotation to the public databases: Sep 10 2015
Software version: 6.4

Assemblies

The following assemblies were included in this annotation run:

Assembly name	Assembly accession	Submitter	Assembly date	Reference/Alternate	Assembly content
Limulus_polyphemus-2.1.2	GCF_000517525.1	Washington University (WashU)	01-03-2014	Reference	1 assembled chromosomes; unplaced scaffolds

Gene and feature statistics

Counts and length of annotated features are provided below for each assembly.

Feature counts

Feature	Limulus_polyphemus-2.1.2
Genes and pseudogenes	22,129
protein-coding	20,409
non-coding	1,270
pseudogenes	450
genes with variants	1,908
mRNAs	23,287
fully-supported	11,424
with > 5% ab initio	8,871
partial	4,232
with filled gap(s)	20

Seikagaku Corporation Exhibit 2004 - Page 208 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Feature	Limulus_polyphemus-2.1.2
known RefSeq (NM_)	0
model RefSeq (XM_)	23,287
Other RNAs	2,066
fully-supported	1,433
with > 5% ab initio	0
partial	0
with filled gap(s)	0
known RefSeq (NR_)	0
model RefSeq (XR_)	1,433
CDSs	23,287
fully-supported	11,424
with > 5% ab initio	9,336
partial	4,231
with major correction(s)	434
known RefSeq (NP_)	0
model RefSeq (XP_)	23,287

Detailed reports

Feature lengths

Feature	Count	Mean length (bp)	Median length (bp)	Min length
Genes	21,679	29,383	18,821	
All transcripts	25,353	1,930	1,467	
mRNA	23,287	2,061	1,570	
misc_RNA	146	1,900	1,572	
tRNA	633	74	73	
lncRNA	1,287	480	304	
Single-exon transcripts	1,897	1,166	921	
coding transcripts (NM_/XM_)	1,897	1,166	921	
CDSs	23,287	1,533	1,119	
Exons	175,047	236	137	
in coding transcripts (NM_/XM_)	170,990	237	138	
in non-coding transcripts (NR_/XR_)	4,752	173	88	
Introns	153,016	4,025	1,948	
in coding transcripts (NM_/XM_)	150,291	4,022	1,946	
in non-coding transcripts (NR_/XR_)	3,392	4,205	2,055	

Transcripts per gene, exons per transcript

Seikagaku Corporation Exhibit 2004 - Page 209 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

	Mean	Median	Min
Number of transcripts per gene	1.14	1	1
Number of exons per transcript	8.2	6	1

Alignment of the annotated proteins to a set of high-quality proteins

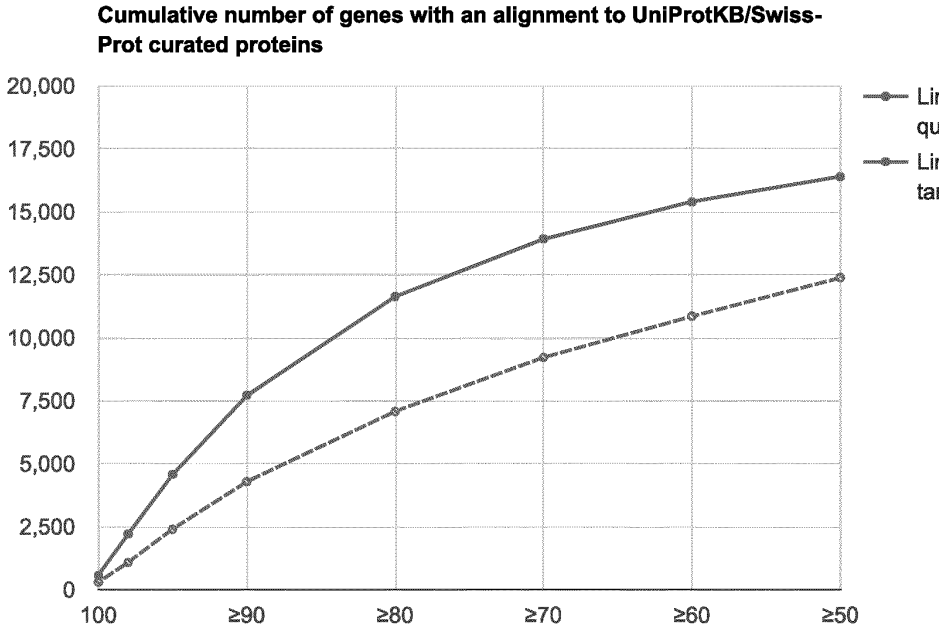
The final set of annotated proteins was searched with BLASTP against the UniProtKB/Swiss-Prot curated proteins, using the annotated proteins as the query and the high-quality proteins as the target. Out of 20409 coding genes, 16391 genes had a protein with an alignment covering 50% or more of the query and 4578 had an alignment covering 95% or more of the query.

Definition of query and target coverage. The query coverage is the percentage of the annotated protein length that is included in the alignment. The target coverage is the percentage of the target length that is included in the alignment.

Below is a cumulative graph displaying the number of genes with alignments above a given query or target coverage threshold. For comparison, corresponding statistics for other organisms annotated by the NCBI eukaryotic annotation pipeline were added to the graph.

Query: annotated proteins
Target: UniProtKB/Swiss-Prot curated proteins

Query & Target



Seikagaku Corporation Exhibit 2004 - Page 210 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Masking of genomic sequence

Transcript and protein alignments are performed on the repeat-masked genome. Below are the percentages of genomic sequence masked by WindowMasker and RepeatMasker for each assembly. RepeatMasker results are only used for organisms for which a comprehensive repeat library is available.

For this annotation run, transcripts and proteins were aligned to the genome masked with WindowMasker only.

Assembly name	Assembly accession	% Masked with RepeatMasker	% Masked with WindowMasker
Limulus_polyphemus-2.1.2	GCF_000517525.1	1.30%	34.26%

Transcript and protein alignments

The annotation pipeline relies heavily on alignments of experimental evidence for gene prediction. Below are the sets of transcripts and proteins that were retrieved from Entrez, aligned to the genome by Splign or ProSplign and passed to Gnomon, NCBI's gene prediction software.

Depending on the other evidence available, long 454 reads (with average length above 250 nt) may be aligned as traditional evidence and reported in the Transcript alignments section or aligned with short reads and reported in the Short read transcript alignments section.

Transcript alignments

Source	Number of sequences retrieved from Entrez	Number (%) of sequences aligned by Splign	Number (%) of sequences passed to Gnomon	Av id
Same-species Genbank	144	136 (94.44%)	123 (85.42%)	
Same-species EST	8,488	6,780 (79.88%)	6,385 (75.22%)	

Short read transcript alignments

The following short reads (RNA-Seq) from the Sequence Read Archive were also used for gene prediction:

Hide alignments statistics, by sample (SAME, SAMN, SAMD, DRS)

Sample Id	Track name	Number of reads	Percent aligned reads	Percent spliced reads	Number of introns
All	Aggregate of all aligned samples	131,549,996	44%	7%	143,975
SAMN00109696	adult and juvenile CNS (Limulus polyphemus, SAMN00109696)	205,260	26%	3%	5,854
SAMN01797697	amoebocyte (Limulus polyphemus, SAMN01797697)	772,309	70%	31%	55,145
SAMN02597592	adult, walking legs (Limulus polyphemus, SAMN02597592)	130,198,888	44%	7%	135,819
SAMN03268704	adult lateral eye (Limulus polyphemus, SAMN03268704)	93,858	48%	27%	23,793
SAMN03268705	adult median eye (Limulus polyphemus, SAMN03268705)	140,730	49%	28%	39,189

Seikagaku Corporation Exhibit 2004 - Page 211 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Sample Id	Track name	Number of reads	Percent aligned reads	Percent spliced reads	Numb of intro
SAMN03268706	adult ventral eye (Limulus polyphemus, SAMN03268706)	138,951	57%	35%	

Show alignments statistics, by run (ERR, SRR, DRR)

Protein alignments

Source	Number of sequences retrieved from Entrez	Number (%) of sequences aligned by ProSplign	Number (%) passed to
Metaseiulus occidentalis model RefSeq (XP_)	11,596	7,342 (63.31%)	7,
Arthropoda GenBank	105,940	69,421 (65.53%)	69,
Same-species GenBank	140	136 (97.14%)	
Ixodes scapularis model RefSeq (XP_)	20,467	12,103 (59.13%)	12
Acyrthosiphon pisum high-quality model RefSeq (XP_)	8,683	5,307 (61.12%)	5
Tribolium castaneum high-quality model RefSeq (XP_)	9,871	6,640 (67.27%)	6,
Drosophila melanogaster known RefSeq (NP_)	29,873	16,864 (56.45%)	16,
Apis mellifera high-quality model RefSeq (XP_)	8,557	5,570 (65.09%)	5,
Homo sapiens known RefSeq (NP_)	39,276	23,611 (60.12%)	23

References

- RefSeq: Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, Murphy MR, O'Leary NA, Pujar S, Rajput B, Rangwala SH, Riddick LD, Shkeda A, Sun H, Tamez P, Tully RE, Wallin C, Webb D, Weber J, Wu W, Dicuccio M, Kitts P, Maglott DR, Murphy TD, Ostell JM. *Nucleic Acids Research* 2014, 42(Database issue):D756-63
- RepeatMasker: Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996–2004. http://www.repeatmasker.org
- WindowMasker: Morgulis A, Gertz EM, Schäffer AA, Agarwala R. *Bioinformatics* 2006, 2:134-41
- Splign: Kapustin Y, Souvorov A, Tatusova T, Lipman D. *Biology Direct* 2008, 3:20

Seikagaku Corporation Exhibit 2004 - Page 212 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

J. Biochem. 109, 150-157 (1991)

Further Studies on Lipopolysaccharide-Sensitive Serine Protease Zymogen (Factor C): Its Isolation from *Limulus polyphemus* Hemocytes and Identification as an Intracellular Zymogen Activated by α-Chymotrypsin, Not by Trypsin[1]

Fuminori Tokunaga,* Hiroshi Nakajima, and Sadaaki Iwanaga*,[2]**

*Department of Biology, Faculty of Science, Kyushu University 33, Higashi-ku, Fukuoka, Fukuoka 812; and **Taiyo Central R&D Institute, Taiyo Fishery Co., Ltd., Chuou-ku, Tokyo 104

Received for publication, August 10, 1990

An intracellular serine protease zymogen, factor C, is an initiator in the hemolymph coagulation system of horseshoe crab. We purified this zymogen from the hemocytes of the American horseshoe crab, *Limulus* (*L.*) *polyphemus*, the objective being to compare its properties with those of the Japanese horseshoe crab, *Tachypleus* (*T.*) *tridentatus*, factor C. The purified zymogen *L.*-factor C showed similar properties to those of *T.*-factor C, in terms of molecular mass (123,000), amino acid composition (1,011 residues), subunit structure (two chains), and antigenicity. Like the zymogen *T.*-factor C, this zymogen was also activated autocatalytically in the presence of bacterial lipopolysaccharide (LPS) and its synthetic lipid A analogue. A most interesting finding is that both protease zymogens are rapidly activated by α-chymotrypsin or rat mast cell chymase, but not by trypsin. The active enzyme factor C̄ showed α-thrombin-like specificity toward synthetic tripeptide substrates. This factor C̄ was also strongly inhibited by an α-thrombin inhibitor, D-Phe-Pro-Arg-chloromethyl ketone. Thus, the enzymatic properties of factor C̄ are similar to those of mammalian α-thrombin. On the other hand, the coagulation cascade system present in the hemocyte lysate was activated when chymotrypsin, free from LPS, was added to the lysate used to detect the endotoxins. The implication of our findings is that the chymotrypsin-catalyzed initiation of the horseshoe crab coagulation system is unique, since all known mammalian coagulation, fibrinolysis and complement systems are initiated by trypsin-like enzymes.

Horseshoe crab hemocytes are highly sensitive to Gram-negative bacterial endotoxins, lipopolysaccharides (LPS) (*1-3*), exposure to LPS resulting in activation of intracellular coagulation systems. In the Japanese horseshoe crab, *Tachypleus* (*T.*) *tridentatus*, the coagulation system consists of at least three intracellular serine protease zymogens, named proclotting enzyme (*4*), factor B (*5*), and factor C (*6-9*), in addition to the fibrinogen-like substance, coagulogen (*10, 11*). These zymogens and coagulogen constitute the cascade pathway responsible for the LPS-mediated activation, as shown in Fig. 1 (*6*). Factor C is an LPS-sensitive protease zymogen consisting of a single chain glycoprotein with a molecular mass of 123,000. During the LPS-mediated activation of factor C, its single chain form is converted to a two chain intermediate form comprising an 80 kDa heavy chain and a 43 kDa light chain, and the latter is subsequently transformed to a 7.9 kDa A chain and a 34 kDa B chain on cleavage of a unique Phe-Ile bond (*6, 7*). This three chain-factor C̄ has the potential to activate zymogen factor B (*5*) and to hydrolyze a synthetic tripeptide substrate, such as Boc-Val-Pro-Arg-NH-Np (*6*). The entire cDNA sequence of this factor C has been determined (manuscript submitted). The deduced amino acid sequence data suggest that the zymogen factor C is a mosaic protein composed of a Cys-rich, an EGF-like, a lectin-like, five β_2GP-I-like, and serine protease domains.

Although much progress has been made on the zymogen, factor C, isolated from *T. tridentatus* hemocytes, little is known of coagulation factors in hemocytes of the American horseshoe crab, *Limulus* (*L.*) *polyphemus*. In 1977, Liu and coworkers (*12, 13*) reported the existence of a proclotting enzyme in *L. polyphemus* hemocytes, and postulated that it is an initiation factor sensitive to LPS and closely associated with the limulus coagulation system. However, Nakamura and Levin (*14, 15*) examined *L. polyphemus* hemocytes and suggested that the lysate contains another LPS-sensitive factor differing from the known proclotting enzyme.

We attempted to shed light on these conflicting reports, and we purified and characterized a factor C-type zymogen from *L. polyphemus* hemocytes and compared its properties with those of the LPS-sensitive factor C from *T. tridentatus* hemocytes. We obtained evidence that the

[1] This work was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan.

[2] To whom correspondence should be addressed.

Abbreviations: LPS, lipopolysaccharide; factor C̄, activated factor C; *T.*, *Tachypleus*; *L.*, *Limulus*; Boc, *N-tert*-butoxycarbonyl; NH-Np, *p*-nitroanilide; NH-Mec, 4-methylcoumaryl-7-amide; OBzl, *O*-benzyl; Z, benzyloxycarbonyl; Tris/HCl/SA buffer, 0.1 M Tris/HCl buffer, pH 8.0, containing 0.5 mg/ml of human serum albumin; MUGB, 4-methylumbelliferyl *p*-guanidinobenzoate; MUTMAC, 4-methylumbelliferyl *p*-(*N,N,N*-trimethylammonium)cinnamate.

Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 213 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

zymogen factor C sensitive to LPS does exist in *Limulus* hemocytes, and that the immunological and biochemical properties of the purified factor C are indistinguishable from those of factor C isolated from *Tachypleus* hemocytes. Moreover, this serine protease zymogen is unique in that it is activated by α-chymotrypsin, not by trypsin.

RESULTS

Characterization of L. polyphemus Factor C—The purity of the final preparation of *L.*-factor C was examined by SDS-PAGE. The purified zymogen factor C gave a single band of 123 kDa under nonreduced conditions, and two bands corresponding to an 80 kDa heavy chain and a 43 kDa light chain after reduction with β-mercaptoethanol, as shown in Supplementary Fig. 4. After activation with LPS, the zymogen factor C was converted to its active form, *i.e.*, factor $\bar{C}$, consisting of three chains of 80, 34, and 8 kDa, respectively, in the same manner as in the case of *T.*-factor C (data not shown). Western blotting performed with rabbit antiserum raised against *T.*-factor $\bar{C}$ indicated that the zymogen *L.*-factor C reacts with anti-*T.*-factor $\bar{C}$ antiserum, three bands corresponding to 123, 80, and 43 kDa being detected under reduced conditions (Supplementary Fig. 4, lane 3). The faint 123 kDa band seems to represent a single chain factor C (*7*). On Ouchterlony analysis using anti-*T.*-factor $\bar{C}$ antiserum, both preparations, *i.e.*, *T.*- and *L.*-factor Cs, formed a fusion type precipitin line (Supplementary Fig. 5).

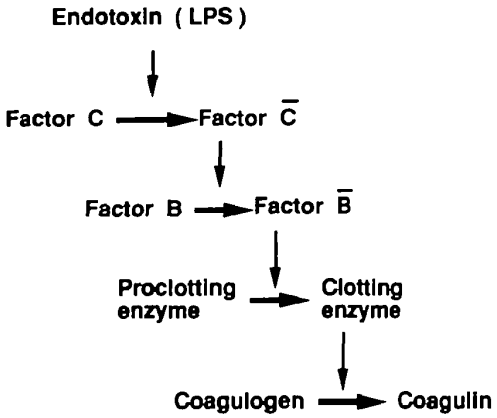


Fig. 1 **Putative mechanism for the coagulation cascade of the horseshoe crab hemocyte lysate.**

Activation of the zymogen *L.*-factor C by LPS and a synthetic *E. coli*- type lipid A analogue was examined by monitoring the generation of amidase activity of factor $\bar{C}$, using Boc-Val-Pro-Arg-NH-Np as a substrate. As shown in Supplementary Fig. 6, both LPS and the lipid A analogue activated the zymogen factor C, the activation reaction being completed within 30 min. Amidase activity did not appear to be generated in the absence of LPS.

Amino Acid Compositions and NH_2-Terminal Sequences of L.-Factor C and Its Chain Subunits—To compare the polypeptide chain structure of *L.*-factor C with that of *T.*-factor C, the purified zymogen factor C and its activated form were reduced and *S*-alkylated, and then the resulting chain subunits were separated by means of the high performance liquid chromatography used for the zymogen *T.*-factor C (*6, 7*). The materials thus obtained were subjected to amino acid and sequence analyses. Supplementary Table II shows the amino acid compositions of the intact *L.*-factor C and its chain subunits. *L.*-factor C was calculated to comprise a total of 1,011 residues per molecule, this number being much the same as that for *T.*-factor C (*6*). Furthermore, the individual amino acid compositions of the two proteins and of their chain subunits were indistinguishable. Like *T.*-factor C (*6*), *L.*-factor C also contained hexosamines, which were present in the acid hydrolyzate.

Supplementary Fig. 7 shows the partial NH_2-terminal

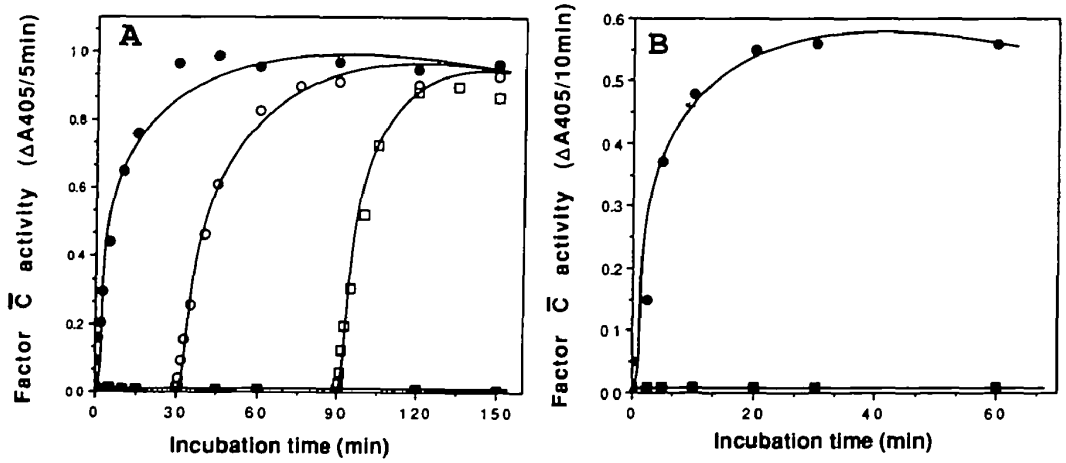


Fig. 2. **α-Chymotrypsin-mediated activation of factor C.** A: The purified *T.*-factor C was incubated, respectively, with α-chymotrypsin (●) and trypsin (■) at an enzyme to substrate weight ratio of 1 : 200. At the indicated times, a sample of the reaction mixture was withdrawn and factor $\bar{C}$ activity was measured as described under "MATERIALS AND METHODS." In other experiments, the reaction mixture containing factor C and trypsin was preincubated, respectively, for 30 (○) and 90 min (□), and then α-chymotrypsin was added after abolition of the trypsin activity with *N*-tosyl-L-lysyl chloromethyl ketone, followed by further incubation, was allowed to continue. The factor $\bar{C}$ activity generated was measured at the indicated times. B: The purified *L.*-factor C (95.6 μg/ml) was treated in the same manner as in A.

Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 214 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

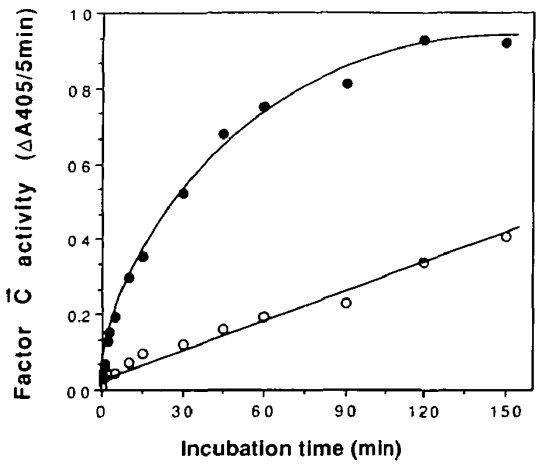


Fig. 3. **Chymase-mediated activation of factor C.** *T.*-factor C was incubated with chymase at enzyme to substrate weight ratios of 1 : 50 (●) and 1 · 200 (○), and then the factor C activity generated was measured as described under "MATERIALS AND METHODS "

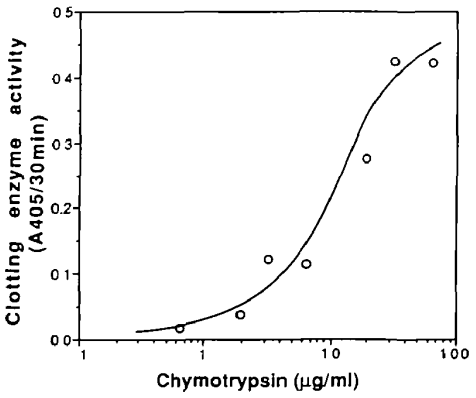


Fig. 4. **Activation of the Limulus coagulation system by α-chymotrypsin.** The toxicolor® reagent containing the hemocyte lysate of *T. tridentatus* was incubated with α-chymotrypsin for 30 min at 37°C and the clotting enzyme activity generated was measured using Boc-Leu-Gly-Arg-NH-Np as substrate For details, see "MATERIALS AND METHODS "

TABLE I. **Kinetic parameters of *T.*-factor C.** The active site-titrated factor C̄ (2 7 nM) was used for kinetic analyses. Kinetic parameters were determined by direct linear plotting. The concentration of *p*-nitroaniline released was estimated by assuming $A^{1\%}_{405nm}$ = 10,000

Substrates				k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}\cdot s^{-1}$)
P4	P3	P2	P1			
Boc-	Val	-Pro	-Arg-NH-Np	391	54	7,240,000
Boc-Asp(OBzl)		-Pro	-Arg-NH-Mec	75	26	2,880,000
Boc-	Gln	-Pro	-Arg-NH-Mec	34	66	514,000
Boc-Asp(OBzl)		-Ala	-Arg-NH-Mec	18	39	489,000
Boc-	Val	-Pro	-Arg-NH-Mec	48	114	418,000
Boc-	Ala	-Pro	-Arg-NH-Mec	72	190	379,000
Boc-	Met	-Ala	-Arg-NH-Mec	27	88	212,000
Boc-	Lys(Z)	-Gly	-Arg-NH-Mec	10	45	227,000
Boc-	Gly	-Ala	-Arg-NH-Mec	29	155	190,000
Boc-Ser(OBzl)		-Ala	-Arg-NH-Mec	26	152	172,000
Boc-	Lys	-Gly	-Arg-NH-Mec	19	126	149,000
Boc-Glu(OBzl)		-Gly	-Arg-NH-Mec	34	275	122,000
Boc-	Val	-Gly	-Arg-NH-Mec	26	250	104,000
Boc-	Ile	-Gly	-Arg-NH-Mec	15	189	80,000
Boc-	Asp	-Pro	-Arg-NH-Mec	17	221	76,000
Boc-	Gly	-Pro	-Arg-NH-Mec	8.5	167	51,000

sequences of the heavy, light and B chains isolated from *L.*-factor C and C̄, in comparison with those of *T.*-factor C. These sequences showed a close resemblance, thus the gross structure of *L.*-factor C must resemble that of known *T.*-factor C, as shown in Supplementary Fig. 8. These results indicate that the zymogen factor C's isolated from hemocytes of American and Japanese horseshoe crabs, respectively, are identical protein entities, both immunologically and biochemically. Thus, the following kinetic studies on zymogen activation and substrate specificities were performed mainly using the purified *T.*-factor C, since it was more readily available.

Time Course for Activation of the Zymogen Factor C with α-Chymotrypsin—Figure 2A shows the time course for activation of *T.*-factor C in the presence of trypsin and α-chymotrypsin. The factor C was rapidly activated at 37°C in the presence of α-chymotrypsin. However, trypsin did not activate factor C at an enzyme to substrate weight ratio of 1 : 200 or even of 1 : 20 (w/w) (data not shown). Such α-chymotrypsin-mediated activation of the *T.*-factor C was also observed for that of the purified *L.*-factor C (Fig. 2B). To determine whether or not trypsin would destroy the zymogen factor C, it was incubated with trypsin at 37°C for 30 and 90 min, and then 20 μg of *N*-tosyl-L-lysyl chloromethyl ketone was added to abolish the trypsin activity. α-Chymotrypsin, at 1 : 200 (w/w), was then added to the reaction mixture. As shown by open circles and squares in Fig. 2A, the zymogen factor C was activated by α-chymotrypsin in the same manner as with α-chymotrypsin alone, confirming again that it is not activated by trypsin. The activation of the factor C by α-chymotrypsin was also examined by SDS-PAGE under reduced conditions (Supplementary Fig. 9). After 30 and 60 min incubation, the factor C-light chain (43 kDa) was converted to a 34-kDa B chain. The factor C-heavy chain was also partially cleaved, because some bands appeared between those of the heavy and B chains upon α-chymotrypsin treatment. As rat mast cell chymase shows a similar specificity to α-chymotrypsin (*17*), we examined the chymase-mediated activation of factor C. As shown in Fig. 3, the zymogen factor C was activated by chymase at enzyme to substrate weight ratios of 1 : 200 and 1 : 50, and a plateau was reached as the same extent as in the case of α-chymotrypsin. The K_m and k_{cat} values of α-chymotrypsin toward the zymogen *T.*-factor C were estimated to be 7.4×10^{-8} M and 5.6×10^{-7} $M^{-1}\cdot s^{-1}$, respectively (data not shown). Other proteases, such as porcine elastase, lysyl endopeptidase from *Achromobacter lyticus* M497-1, *Staphylococcus aureus* V8 protease, subtilisin, and papain, did not activate the zymogen factor C (data not shown).

Activation of the Horseshoe Crab Coagulation System by α-Chymotrypsin—Figure 4 shows the α-chymotrypsin-mediated activation of horseshoe crab coagulation with the hemocyte lysate. As expected, the amidase activity generated due to the activation of proclotting enzyme (*4*) increased in a dose-dependent manner. When α-chymo-

Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 215 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

trypsin treated with Z-Gly-Gly-Phe-chloromethyl ketone for 15 min at 37°C or the enzyme treated at 70°C for 1 h was used, the amidase activity generated in the hemocyte lysate decreased to 73.4 and 89.5%, respectively (data not shown). These results indicate that activation of the coagulation system was not induced by LPS contaminating the α-chymotrypsin sample, but rather that it was catalyzed by α-chymotrypsin.

Substrate Specificity of Factor C̄—Since the purified factor C was activated autocatalytically in the presence of LPS through the cleavage of a Phe-Ile linkage located in the NH_2-terminal portion of the light chain (*7*), the possibility that factor C itself might have intrinsic α-chymotrypsin-like activity would have to be considered. We examined in detail the substrate specificity of factor C̄ toward various synthetic peptidyl substrates, including substrates for α-chymotrypsin. The concentration of factor C̄ was first determined by active-site titration with 4-methylumbelliferyl *p*-guanidinobenzoate (MUGB), which is used as a titrant for trypsin-type proteases. However, the rate of deacylation of the reagent was too rapid to estimate the enzyme concentration, as shown in Supplementary Fig. 10A. Therefore, 4-methylumbelliferyl *p*-(*N,N,N*-trimethylammonium)cinnamate (MUTMAC), synthesized as a suitable titrant for α-chymotrypsin (*18*), was used for the titration of factor C̄. MUTMAC proved to be an appropriate titrant for factor C̄, as deduced from the burst and stable saturation curve for the titration, as shown in Supplementary Fig. 10B. As this means that factor C probably has an intrinsic chymotrypsin-like activity, this titrant was used to determine the concentration of active factor C̄.

Table I shows the kinetic constants of factor C̄ toward various peptidyl substrates. The most effective substrate for factor C̄ was Boc-Val-Pro-Arg-NH-Np. It has a high k_{cat}/K_m value, 7,240,000, which is 17-fold higher than that of Boc-Val-Pro-Arg-NH-Mec. Among 75 peptidyl-NH-Mec substrates (*14*), Boc-Asp(OBzl)-Pro-Arg-NH-Mec was also a good substrate factor C̄, thereby indicating that the enzyme effectively cleaves substrates with a Pro residue at the P2 site. The NH-Mec substrates for α-chymotrypsin, such as succinyl-Ala-Ala-Pro-Phe-NH-Mec and succinyl-Leu-Leu-Val-Tyr-NH-Mec, were not hydrolyzed by factor C̄ (data not shown).

Inhibition of Factor C̄ by Natural and Synthetic Inhibitors—To further elucidate the enzymatic properties of factor C, the inhibitory effects of some natural or synthetic inhibitors on factor C̄ activity were examined. Natural inhibitors, such as antithrombin III-heparin complex, and synthetic inhibitors, such as di*iso*propylfluorophosphate, benzamidine, and leupeptin, are known to inhibit factor C̄ activity (*6*). Other natural inhibitors tested were aprotinin, soybean trypsin inhibitor, potato II inhibitor, C1 inhibitor, corn inhibitor, *Streptomyces* subtilisin inhibitor, hirudin, ovomucoid, and ovoinhibitor. However, these inhibitors had little effect on the factor C̄ activity, even in the presence of an excess molar ratio to the enzyme. Among the synthetic inhibitors, *N*-tosyl-L-lysyl chloromethyl ketone, chymostatin, Z-Gly-Gly-Phe-chloromethyl ketone, and H-Ala-Ala-Phe-chloromethyl ketone also did not inhibit factor C̄ activity, at a final concentration of 1 mM. On the other hand, D-Phe-Pro-Arg-chloromethyl ketone, an α-thrombin inhibitor, had a potent inhibitor effect on the factor C̄ activity. The inhibition constant of this inhibitor towards factor C̄ was determined by the method of Walker *et al.* (*19*), and the K_i value was estimated to be 7.4×10^{-6} M and k_i to be 0.69 min^{-1}(data not shown).

DISCUSSION

A factor C, which is indistinguishable from those of the known *T.*-factor C in terms of immunological and biochemical properties, was proved to be present in the hemocytes of *L. polyphemus*. Both factor C's contain two molecular species with, *i.e.*, single chain and a two-chain forms. Both molecules have the same molecular mass of 123,000, but the latter consists of a heavy chain of $M_r=80{,}000$ and a light chain of $M_r=43{,}000$ held together by a disulfide linkage(s). The major nicked form may be produced by a trace amount of contaminating of LPS during the purification procedures, as experienced in the study on *T.*-factor C (*6*). The structural properties of the *L.*-factor C molecule, including the entire amino acid composition and the partial NH_2-terminal sequences of their chain subunits, show similarities with those of *T.*-factor C (*6, 7*). They also exhibit common antigenicity, as revealed with the immunoblotting method and the Ouchterlony test (Supplementary Figs. 4 and 5). Like the zymogen *T.*-factor C, the purified *L.*-factor C is rapidly activated in the presence of *E. coli* LPS and synthetic lipid A analogue. These observations strongly indicate that both factor C's are the same protease zymogen and function as an initiator of the coagulation cascade in the horseshoe crab hemocytes. These findings also lead to the idea that the hemocytes of *L. polyphemus* probably contain a coagulation system similar to that established for the *T. tridentatus* hemocytes (*5-7*).

The present results are the first evidence that both the purified *L.*- and *T.*-factor C's are activated by α-chymotrypsin and not by trypsin. All the serine protease zymogens associated with mammalian coagulation, fibrinolysis, and complement systems are activated by trypsin-like enzymes. Thus, the horseshoe crab factor C seems to be a novel type of serine protease zymogen susceptible to α-chymotrypsin. The necessity of cleavege of the Phe-Ile bond (*7*) implies that the zymogen factor C cannot be activated by the horseshoe crab trypsin-like proteases (*4, 5*). Thus, positive feedback activation of the clotting system, which occurs in common in the mammalian coagulation cascade reaction, probably does not occur in horseshoe crabs, only the LPS-associated factor C appearing to be the initiator of this cascade reaction.

For the LPS-mediated activation of the zymogen factor C, it is now clear that no other protease is required for its conversion to its active form, factor C̄, and only LPS-bound factor C is activated, probably through an intermolecular interaction between the two molecules (*8*), as is the case with activation of the complement Clr (*20*). The substrate specificity of factor C̄ shows a close similarity to that of α-thrombin. As shown in Table I, the most reactive substrate for factor C̄ is Boc-Val-Pro-Arg-NH-Np, with an extremely high $k_{cat}/K_m=7{,}240{,}000$. This value is comparable to that of the sensitive fluorogenic substrate, Boc-Asp(OBzl)-Pro-Arg-NH-Mec, for bovine α-thrombin (*14*). Thus, apolar and non-aromatic residues such as Pro and Ala at the P2 site are the most favorable, as for α-thrombin (*16*). For the P3 site, bulky groups such as Val, Asp(OBzl), and Gln probably serve, as judged from their K_m values

Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 216 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

(Table I). Such a structural basis for a peptide substrate for factor C is also supported by the observation that D-Phe-Pro-Arg-chloromethyl ketone, which is sensitive to α-thrombin strongly inhibits factor $\bar{C}$ activity, the K_i value being 7.4×10^{-6} M. The entire cDNA sequence and deduced amino acid sequence of *T.*-factor C have been determined (manuscript submitted). Then, the factor C-derived B chain corresponding to a serine protease domain of this zymogen is the most homologous with human α-thrombin (38.5% identity) among various known mammalian serine proteases. Presumably, the serine protease domain of *T.*-factor C is closely related with α-thrombin, in terms of molecular structure and substrate specificity.

Although the active factor $\bar{C}$ shows the Arg-Xaa specificity toward synthetic peptide substrates, cleavage of the Phe-Ile bond by LPS-bound factor $\bar{C}$ is not unreasonable, because the residue preceding Phe is Pro (*7*). The reactive site of α_1-antitrypsin has been identified as Met-Ser with Met proceded by Pro (*21*). Moreover, the reactive site of heparin cofactor II is a Leu-Ser bond that is cleaved by α-thrombin. The rapid hydrolysis of MUTMAC by the LPS-bound factor C indicates that it should have intrinsic chymotrypsin-like activity. It seems likely that structural properties of horseshoe crab factor C beyond the specific cleavage site residues play an important role in the LPS-mediated activation of this zymogen. Further detailed studies on the molecular interaction of LPS and factor C are underway.

We wish to express our thanks to Drs. T Nakamura and S. Kawabata for the helpful discussions, T. Miura for the technical assistance, N Ueno for the secretarial services, and M. Ohara for the comments on the manuscript.

REFERENCES

1. Bang, F B (1956) *Bull Johns Hopkins Hosp.* **98**, 325-351
2. Levin, J. & Bang, F.B. (1964) *Bull. Johns Hopkins Hosp.* **115**, 265-274
3. Mürer, E.H., Levin, J., & Holme, R. (1975) *J. Cell. Physiol.* **86**, 533-542
4. Nakamura, T., Morita, T., & Iwanaga, S (1985) *J. Biochem.* **97**, 1561-1574
5. Nakamura, T., Horiuchi, T., Morita, T., & Iwanaga, S. (1986) *J. Biochem.* **99**, 847-857
6. Nakamura, T , Morita, T., & Iwanaga, S. (1986) *Eur. J. Biochem* **154**, 511-521
7. Tokunaga, F , Miyata, T , Nakamura, T., Morita, T., Kuma, K., Miyata, T., & Iwanaga, S. (1987) *Eur J Biochem.* **167**, 405-416
8. Nakamura, T., Tokunaga, F , Morita, T , & Iwanaga, S. (1988) *J. Biochem.* **103**, 370-374
9. Nakamura, T., Tokunaga, F., Morita, T., Iwanaga, S , Kusumoto, S., Shiba, T., Kobayashi, T., & Inoue, K. (1987) *Eur. J. Biochem.* **176**, 89-94
10. Tai, J.Y., Seid, R.C., Jr., Huhn, R.D , & Liu, T.-Y. (1977) *J. Biol. Chem.* **252**, 4773-4776
11 Miyata, T., Hiranaga, M., Umezu, M., & Iwanaga, S. (1984) *J Biol Chem.* **259**, 8924-8933
12 Tai, J.Y & Liu, T.-Y (1977) *J. Biol. Chem.* **252**, 2178-2181
13. Liang, S.-M., Liang, C.-M , & Liu, T -Y. (1981) *J Biol. Chem.* **256**, 4968-4972
14 Nakamura, S. & Levin, J (1982) *Biochem. Biophys Res. Commun.* **108**, 1619-1623
15 Nakamura, S. & Levin, J. (1982) *Biochim. Biophys Acta* **707**, 217-225
16 Kawabata, S., Miura, T., Morita, T., Kato, H., Fujikawa, K , Iwanaga, S., Takada, K., Kimura, T., & Sakakibara, S. (1988) *Eur J Biochem.* **172**, 17-25
17. Woodbury, R.G., Everitt, M.T., & Neurath, H. (1981) *Methods Enzymol.* **80**, 588-597
18. Jameson, G.W., Roberts, D.V , Adams, R W , Kyle, W.S.A., & Elmore, D.T. (1973) *Biochem. J.* **131**, 107-117
19 Walker, B , Wikstrom, P , & Shaw, E (1985) *Biochem. J* **230**, 645-650
20 Kashihara, Y., Odai, H , Takahashi, K., Nagasawa, S., & Koyama, J. (1985) *J. Biochem* **97**, 365-372
21. Johnson, D. & Travis, J (1978) *J. Biol. Chem.* **253**, 7142-7144
22 Laemmli, U.K. (1970) *Nature* **227**, 680-685
23. Eisenthal, R. & Cornish-Bowden, A. (1974) *Biochem J.* **139**, 715-720

Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 217 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Supplemental Materials

MATERIALS AND METHODS

Materials ——— American horseshoe crabs (Limulus polyphemus) were obtained from Taiyo Fishery Co., Ltd., Tokyo. T. tridentatus factor C (two chain form) from hemocytes of Japanese horseshoe crabs was highly purified, as described (6). The active enzyme, factor C, was prepared from the isolated factor C by treatment with LPS (6). The LPS (Escherichia coli 0111:B4), obtained from List Biochemical Laboratories, Inc., CA, was heated at 70 °C for 1 h before use. Synthetic E. coli type lipid A analog, LA-15-PP, and aprotinin were products of Daiichi Pharm., Co., Ltd., Tokyo, and Mochida Pharm., Co., Ltd., Tokyo, respectively. Bovine pancreatic α-chymotrypsin treated with N-tosyl-L-lysyl chloromethyl ketone and trypsin treated with N-tosyl-L-phenylalanyl chloromethyl ketone were obtained from Worthington Biochemical Corp., Freehold, NJ. Boc-Val-Pro-Arg-NH-Np and an endotoxin detection kit, Toxicolor^R LS-1 set, which contains the T. tridentatus hemocyte lysate, were generous gifts from S. Tanaka, Seikagaku Kogyo Co., Ltd., Tokyo. Rat mast cell chymotrypsin-like enzyme, chymase (RMCP I), was kindly provided by Dr. N. Katunuma, Institute of Enzyme Research, The University of Tokushima. Fluorogenic tripeptide substrates (14) were obtained from Peptide Institute Inc., Osaka. 4-Methylumbelliferyl p-guanidinobenzoate (MUGB) 4-methylumbelliferyl p-(N N N-trimethylammonium) cinnamate (MUTMAC), and soybean trypsin inhibitor were purchased from Sigma Chemical, Co., St. Louis, MO. H-Ala-Ala-Phe-chloromethyl ketone and Z-Gly-Gly-Phe-chloromethyl ketone were from Bachem, Bubendorf, Switzerland, and D-Phe-Pro-Arg-chloromethyl ketone was from Calbiochem, La Jolla, CA Sepharose CL-6B and CM-Sepharose CL-6B were purchased from Pharmacia LKB Biotechnology, Uppsala. All other chemicals were of the highest quality commercially available.

Factor C assay ——— The amidase activity of factor C generated after activation with LPS was measured using Boc-Val-Pro-Arg-NH-Np as a substrate, as described (6). The reaction mixture, comprising 10-20 µl of sample in Tris/HCl/SA buffer in a total volume of 200 µl, was preincubated at 37 °C for 10 min in the presence or absence of 4 µl of LPS (10 µg/ml). Then, 50 µl of 2 mM Boc-Val-Pro-Arg-NH-Np was added and the mixture was incubated further for 10 min. Finally, 0.8 ml of 0.6 M acetic acid was added to terminate the reaction and the absorbance at 405 nm was read. One unit of enzyme activity was defined as 1 µmol of p-nitroaniline liberated per min. The amidase activity of factor C (active form) was measured as described for factor C in the absence of LPS.

SDS-polyacrylamide gel electrophoresis and Western blotting ——— SDS-PAGE was performed according to the method of Laemmli (22). Western blotting was performed using rabbit antiserum raised against T.-factor C, as described (7). For Ouchterlony analysis, a 1 % agarose gel dissolved in 20 mM Tris-HCl, pH 8.0, and containing 0.15 M NaCl was used.

Reduction and S-alkylation ——— The sample was reduced and S-pyridylethylated and each chain was separated by reversed phase liquid chromatography on a TSK phenyl 5PWRP column (4.6 x 75 mm), as described (7).

Amino acid and sequence analyses ——— Samples were hydrolyzed in 5.7 M HCl at 110 °C for 24, 48 and 72 h or 3 M mercaptoethanesulfonic acid for 24 h in evacuated and sealed tubes. After evaporation, the hydrolyzates were analyzed using a Hitachi 835 automatic analyzer. Amino acid sequence determination was carried out with a gas-phase sequencer (Applied Biosystems 477A) and a 120A-phenylthiohydantoin analyzer on-line system (7).

Activation of factor C by trypsin, α-chymotrypsin and chymase ——— Factor C (206 µg/ml) was incubated at 37 °C with 1/200 (w/w) of trypsin, α-chymotrypsin or chymase, in a total volume of 350 µl of 0.1 M Tris/HCl buffer, pH 8.0. At the indicated times, a 10 µl sample was taken and mixed with 50 µg/ml of aprotinin for α-chymotrypsin and trypsin, or 50 µg/ml of soybean trypsin inhibitor for chymase. These protease inhibitors were dissolved in a total 200 µl of Tris/HCl/SA buffer. Each mixture was incubated for a further 5 min to inhibit the protease activity and finally the factor C activity generated was measured.

Activation of the horseshoe crab coagulation system by α-chymotrypsin ——— α-Chymotrypsin-mediated activation of horseshoe crab coagulation system was measured using Toxicolor^R reagent, LS-1 set, obtained from Seikagaku Kogyo, Ltd., Tokyo, according to the manufacturer's manual. The LS-1 set contained Tachypleus hemocyte lysate and Boc-Leu-Gly-Arg-NH-Np, the latter of which is an appropriate substrate for the clotting enzyme generated during activation of the coagulation system. The LS-1 set (100 µl) and various concentrations of α-chymotrypsin dissolved in 100 µl of 20 mM Tris/HCl buffer, pH 8.0, containing 0.15 M NaCl, were mixed and incubated for 30 min at 37 °C, and then 0.8 ml of 0.6 M acetic acid was added and the absorbance at 405 nm was read.

Kinetic analysis ——— The factor C activity was measured at 37 °C in 50 mM Tris/HCl buffer, pH 8.0, containing 0.15 M NaCl and 0.1 mg/ml of bovine serum albumin (BSA), with various fluorogenic peptide substrates. The initial rate of hydrolysis was measured with a Hitachi 650 fluorometer with excitation at 380 nm and emission at 440 nm at 37 °C, as described in detail elsewhere (14). Substrate concentrations of 1 µM to 2 mM and 2.7 nM factor C were used. Kinetic parameters were determined by direct linear plotting (23). The concentration of active factor C was determined with MUTMAC from the initial burst of 4-methylumbelliferone in 0.1 M sodium barbiturate buffer, pH 8.3, containing 0.1 mg/ml of BSA at 25 °C (18). Kinetics analysis for irreversible inhibitors, such as D-Phe-Pro-Arg-chloromethyl ketone, was performed by the method of Walker et al. (19).

Inhibitory effects of natural and synthetic inhibitors on factor C ——— The screening of various inhibitors was performed with an inhibitor concentration of 0.2 mM and a factor C concentration of 10 nM in Tris/HCl/SA buffer. Each mixture was incubated at 37 °C for 10 min, in a total volume of 200 µl, and then the remaining factor C activity was measured.

RESULTS

Purification of L. polyphemus factor C ——— Purification of L.-factor C was carried out under sterilized conditions according to the method of Nakamura et al. (6). The lysate (595 ml) was first applied to a dextran-sulfate Sepharose CL-6B column (2.6 x 24.0 cm). Stepwise elution was performed with NaCl concentrations of from 50 mM, 0.25 M, 0.5 M and 2 M, as shown in Supplementary Fig. 1. Factor C in the 0.5 M NaCl fractions was collected, pooled and concentrated by Diaflo ultrafiltration. The concentrated fraction was then applied to a Sepharose CL-6B column (2.6 x 141 cm), equilibrated with 20 mM acetate buffer, pH 5.0, containing 0.1 M NaCl (Supplementary Fig. 2). The pooled factor C fraction was finally applied to a CM-Sepharose CL-6B column (2.0 x 24.0 cm), equilibrated with the same buffer as used in Supplementary Fig. 2 and eluted with a linear gradient of NaCl (0.1-0.35 M) in 20 mM acetate buffer, pH 5.0 (Supplementary Fig. 3). A summary of the purification of factor C is shown in Supplementary Table I. The yield of purified factor C was 5.2 mg from 80 g hemocytes (wet weight), 14.7-fold purification from the dextran-sulfate Sepharose CL-6B fraction (Supplementary Fig. 1) being achieved (6).

Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 218 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Supplementary Table I

Summary of the purification procedure for L. polyphemus factor C.

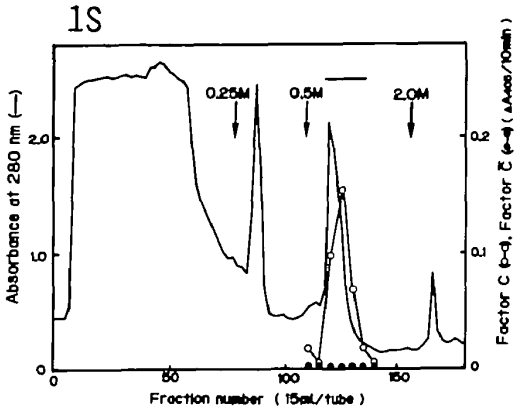
Step	Volume	Total protein	Total activity	Specific activity	Purification	Yield
	ml	mg	units	units/mg	-fold	%
Hemocyte lysate	298	—	—	—	—	—
Dextran-sulfate-Sepharose CL-6B	138	140	397	2.8	1	100
Sepharose CL-6B	90	17.3	421	24.3	8.7	106
CM-Sepharose CL-6B	130	5.2	214	41.2	14.7	54

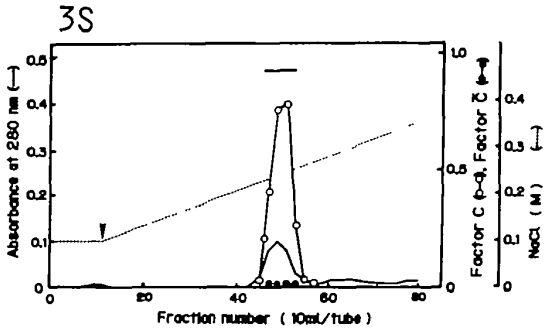
Supplementary Table II

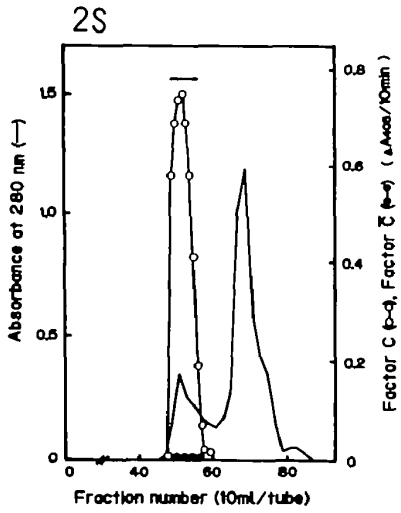
Amino acid compositions of L. polyphemus factor C and its S-alkylated derivatives[a].

Amino acid	Factor C			Factor $\bar{C}$	
	Whole	H chain	L chain	A chain	B chain
	Residues/molecule				
Asp	101.0 (101)	69.4 (67)	39.4 (34)	4.8 (5)	34.6 (29)
Thr	53.6 (59)	36.7 (37)	20.5 (22)	2.2 (2)	18.7 (20)
Ser	88.0 (82)	59.7 (55)	31.5 (27)	11.2 (13)	25.1 (14)
Glu	92.1 (89)	60.7 (58)	34.1 (31)	5.8 (5)	30.8 (26)
Pro	60.8 (69)	47.4 (52)	23.7 (17)	6.9 (6)	20.1 (11)
Gly	89.9 (88)	55.4 (57)	32.6 (29)	7.7 (7)	27.6 (22)
Ala	57.4 (56)	34.7 (35)	22.1 (21)	3.0 (3)	18.3 (18)
Val	61.3 (64)	44.7 (42)	23.8 (22)	3.9 (4)	20.4 (18)
Met	9.3 (10)	6.3 (7)	3.5 (3)	— (0)	3.2 (3)
Ile	39.5 (41)	22.9 (23)	16.0 (18)	2.2 (2)	13.3 (16)
Leu	69.0 (71)	45.4 (47)	25.4 (24)	3.9 (4)	22.2 (20)
Tyr	36.3 (37)	22.1 (22)	14.6 (15)	4.2 (4)	11.4 (11)
Phe	29.4 (28)	20.4 (20)	12.3 (8)	1.9 (1)	14.6 (7)
Lys	54.7 (50)	37.8 (38)	15.9 (12)	3.4 (2)	15.0 (10)
His	17.7 (17)	13.1 (12)	5.3 (5)	1.1 (1)	5.1 (4)
Trp	16.0[b] (23)	11.9 (12)	8.9 (11)	1.6 (1)	8.0 (10)
Pe-Cys	n.d.[b] (56)	37.5 (45)	13.5 (11)	3.4 (4)	12.3 (7)
Arg	44.4 (53)	25.6 (34)	15.8 (19)	6.8 (8)	11.8 (11)
Total	920.4 (994)	651.7 (665)	358.9 (329)	74.0 (72)	312.5 (257)
$GlcNH_2$	+	+	+	-	+
$GalNH_2$	+	+	+	-	+

a: Numbers in parentheses are taken from the sequence data for T.-factor C (manuscript summitted).
b: not determined.







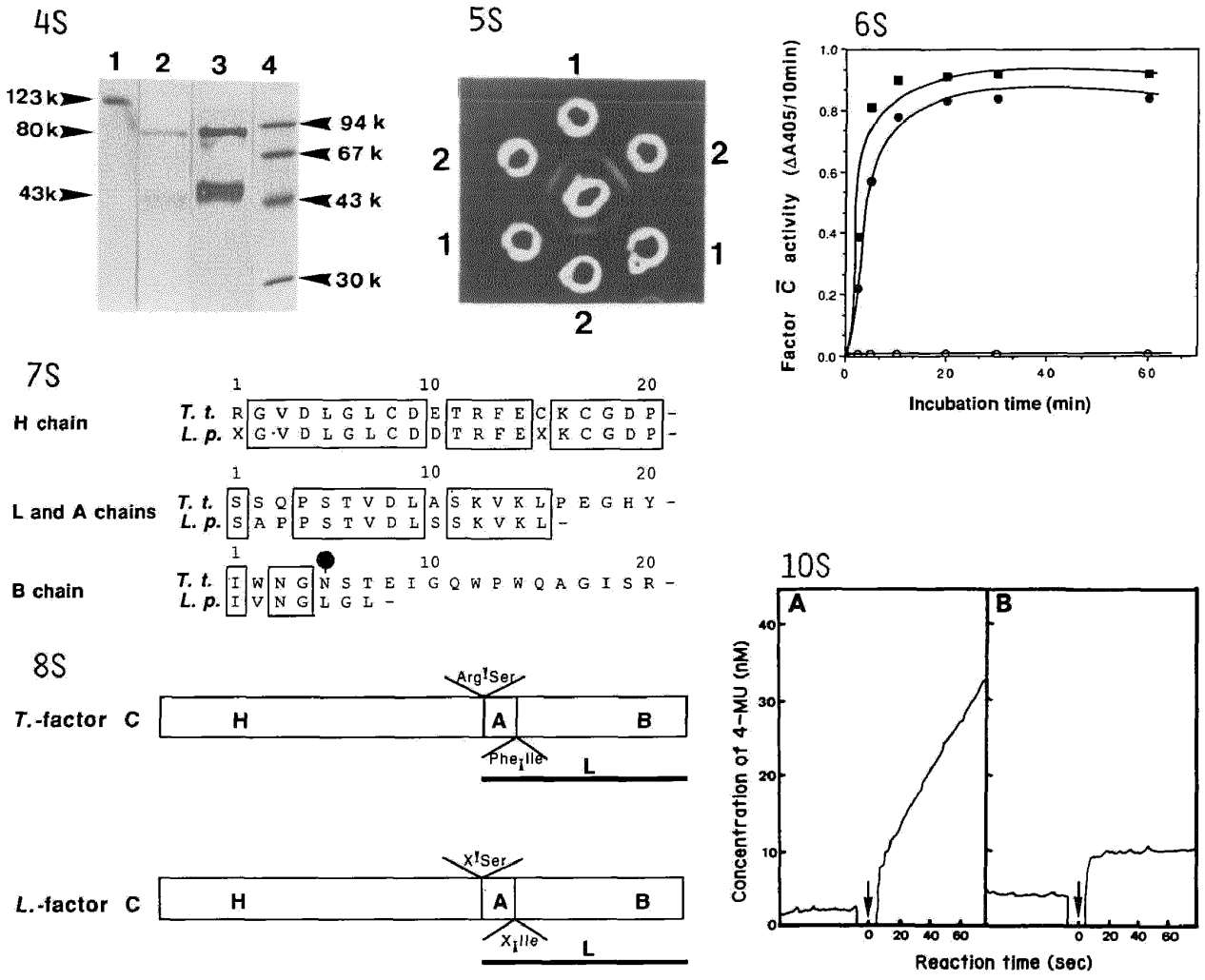
Supplementary Fig. 1. Dextran-sulfate Sepharose CL-6B chromatography of the L. polyphemus hemocyte lysate. Details of the procedure are given under Methods. The fractions indicated by a solid bar were pooled and concentrated. The activities of factor C (○) and its active form, factor $\bar{C}$ (●), were assayed, using Boc-Val-Pro-Arg-NH-Np as a substrate. the absorbance at 280 nm (——) was measured.

Supplementary Fig. 2. Sepharose CL-6B gel filtration of the L.-factor C fraction obtained from a dextran-sulfate Sepharose CL-6B column. The zymogen factor C (○) and the activity of factor $\bar{C}$ (●) were assayed, using Boc-Val-Pro-Arg-NH-Np as a substrate. The absorbance at 280 nm (——) was measured.

Supplementary Fig. 3. CM-Sepharose CL-6B chromatography of the L.-factor C fraction obtained from a Sepharose CL-6B column. The zymogen factor C (○) and the activity of factor $\bar{C}$ (●) were assayed, using Boc-Val-Pro-Arg-NH-Np as a substrate. The absorbance at 280 nm (——) was measured and gradient elution (· ·) with NaCl was started at the position indicated by the arrowhead.

Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 219 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023



Supplementary Fig. 4. SDS-PAGE and Western blotting of the purified L.-factor C. Ten μg of L.-factor C was subjected to SDS-PAGE in the absence (lane 1) and presence (lane 2) of 2-mercaptoethanol. Western blotting (lane 3) of the reduced L.-factor C was also performed using rabbit anti-T.-factor C serum as the 1st antibody, and the positions of co-electrophoresed standards of known molecular weight (lane 4) are indicated.

Supplementary Fig. 5. Immunodiffusion of the purified L.-factor C (1) and T.-factor C (2) against rabbit anti-T.-factor C antiserum (center well).

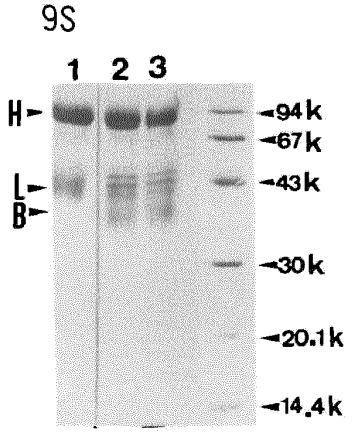
Supplementary Fig. 6. Time course of activation of the L.-factor C in the presence of LPS and a lipid A analogue. The purified L.-factor C (95.6 μg/ml) was incubated at 37 °C in the presence (■) and absence (○) of LPS (10 μg/ml). E. coli type-synthetic lipid A analogue, LA-15-PP (● , 10 μg/ml), was also used. At the indicated times, 20 μl aliquots were withdrawn and assayed for factor C activity.

Supplementary Fig. 7. Comparison of the NH_2-terminal sequences of the heavy, light and B chains derived from each of L.-factor C and T.-factor C. Residues identical in the two factor C's are boxed. The solid circle indicates the carbohydrate attachment site (7).

Supplementary Fig. 8. Gross structure of T.- and L.- factor C's. H, A, B and L denote the heavy, A, B and light chains of factor C. The arrowheads and solid bars indicate the cleavage sites and the locations of the light chains of the factor C's, respectively.

Supplementary Fig. 9. SDS-PAGE of T.-factor C activated by α-chymotrypsin. Samples of the zymogen factor C (lane 1) and the factor C treated with α-chymotrypsin (weight ratio of enzyme/substrate = 1/200) for 30 min (lane 2) and 60 min (lane 3) were subjected to SDS-PAGE. H, L and B denote the heavy, light and B chains of the factor C, respectively.

Supplementary Fig. 10. Active site titration of T.-factor C. A, 4-methylumbelliferyl p-guanidinobenzoate (MUGB) and B, 4-methylumbelliferyl p-(N N N-trimethylammonium) cinnamate (MUTMAC) were used as titrants, being added, respectively, to the reaction mixture containing T.-factor C at the times indicated by arrows after preincubation for 60 sec.



Downloaded from https://academic.oup.com/jb/article-abstract/109/1/150/774612 by University of Winnipeg user on 22 January 2019

Seikagaku Corporation Exhibit 2004 - Page 220 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

US005716834A

United States Patent [19]

Ding et al.

[11] Patent Number: 5,716,834

[45] Date of Patent: Feb. 10, 1998

[54] **CLONED FACTOR C CDNA OF THE SINGAPORE HORSESHOE CRAB, *CARCINOSCORPIUS ROTUNDICAUDA* AND PURIFICATION OF FACTOR C PROENZYME**

[75] Inventors: **Jeak Ling Ding**; **Bow Ho**, both of Singapore, Singapore

[73] Assignee: **National University of Singapore**, Singapore

[21] Appl. No.: **296,014**

[22] Filed: **Aug. 19, 1994**

[51] **Int. Cl.**[6] **C12N 9/50**; C12N 15/55; C12N 15/63

[52] **U.S. Cl.** **435/219**; 435/240.2; 435/252.33; 435/254.11; 435/320.1; 536/23.2

[58] **Field of Search** 435/219, 252.3, 435/320.1, 240.2, 252.33, 254.11; 536/23.2

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,322,217	3/1982	Dikeman	435/7.21
5,082,782	1/1992	Gibson	435/240.2

OTHER PUBLICATIONS

Ho (1983) *Microbios Letters* 24:81–84.

Ho et al. (1985) *Proc. 1st Intl. Cong. Singapore Soc. Microbiol*, pp.664–669.

Kim et al. (1987) Singapore Society for Microbiology, 1987 Annual Scientific Meeting, p. 21.

Ding et al. (1988) *Cytobios* 55:147–154.

Kim et al. (1988) 7th FAOB Symposium, POS–F–01.

Navas et al. (1988) 5 FAOB Congress, MO:19.

Yeo et al. (1989) Second SSM International Congress for Microbiology, BE8.

Navas et al. (1990) *Biochemistry International* 21(5):805–813.

B. Ho et al. (1993) *Biochemistry and Molecular Biology International* 29(4):687–694.

Ding et al (1993) *Biochimica et Biophysica Acta* 1202:149–156.

Ding et al. (1993) *Cytobios* 75:21–32.

Muta et al. (1991) *The Journal of Biological Chemistry* 266(10):6554–6561.

Navas III et al., Proc. 5th FAOB Congress, MO:19, pp. 199 (Aug. 13–18, 1988).

Roopashree et al., Microbiology in the Nineties, pp. 380 (1994).

Ding et al., Biochimica et Biophysica Acta, 1202:149–156 (Sep. 1993).

Ding et al., Essays in Zoology, pp. 337–341 (1990).

Primary Examiner—Dian C. Jacobson

Attorney, Agent, or Firm—Birch, Stewart, Kolasch & Birch, LLP

[57] **ABSTRACT**

Full-length and deletion subclones of cDNAs for Factor C of *Carcinoscorpius rotundicauda* are provided. These cDNAs have been cloned into λgt 22 and pGEM 11Zf(+). Further manipulations of the 5' and 3' ends of these cDNAs have been carried out, and these cDNAs have been further subcloned into other expression vectors such as pGEMEX–1, pET 3b, and the yeast shuttle vectors YEpsec 1 and pEMBLyex 4, and pPIC 9 and pHIL D2. Also provided are host cells transformed with expression vectors containing DNA molecules encoding proteins having Factor C-like enzymatic activity, methods of producing such proteins, methods for purifying Factor C zymogens, and methods for protecting Factor C zymogens from autoactivation by Gram negative bacterial endotoxin while the proenzyme is being purified and/or processed from amoebocyte lysates or from recombinant clones, or during storage or subsequent handling. This protection is afforded by the addition of 5–30% Me_2SO, which reversibly inhibits the Factor C zymogen.

12 Claims, 45 Drawing Sheets

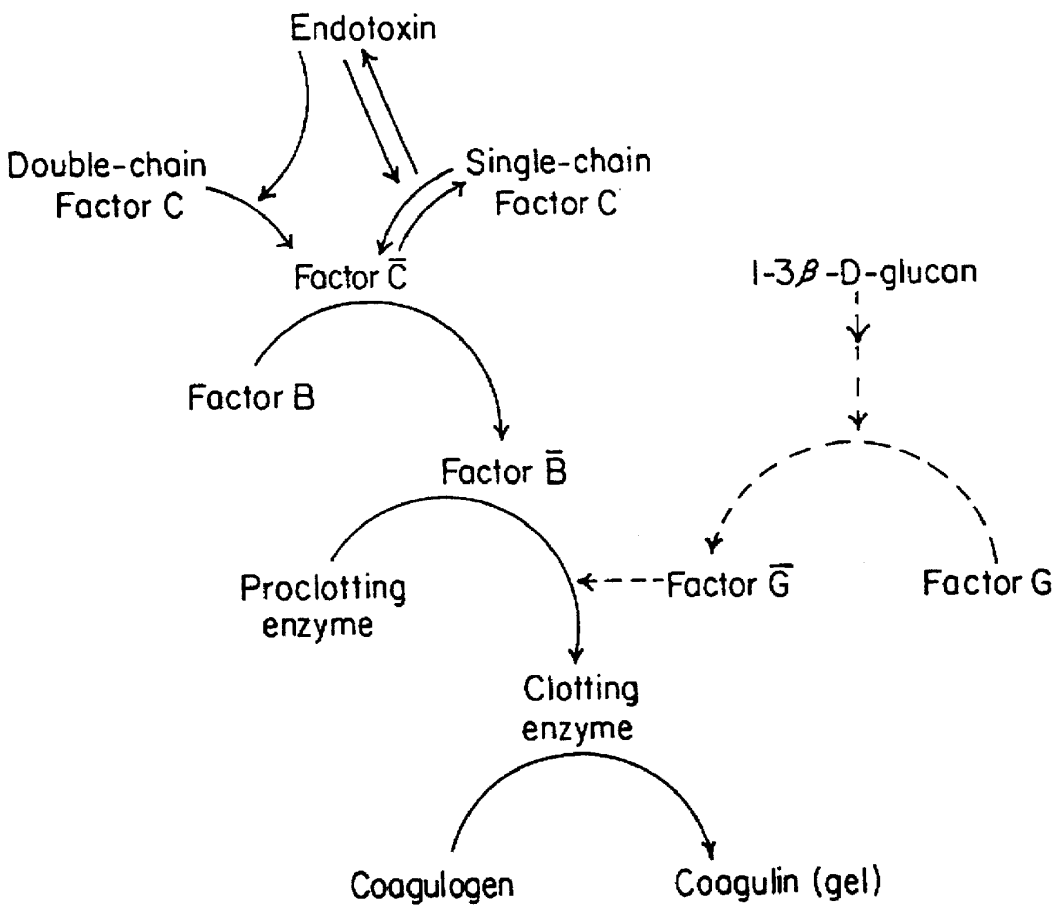

Endotoxin
Double-chain Factor C
Single-chain Factor C
Factor C̄
1-3β-D-glucan
Factor B
Factor B̄
Proclotting enzyme
Factor Ḡ
Factor G
Clotting enzyme
Coagulogen
Coagulin (gel)

FIG. 1

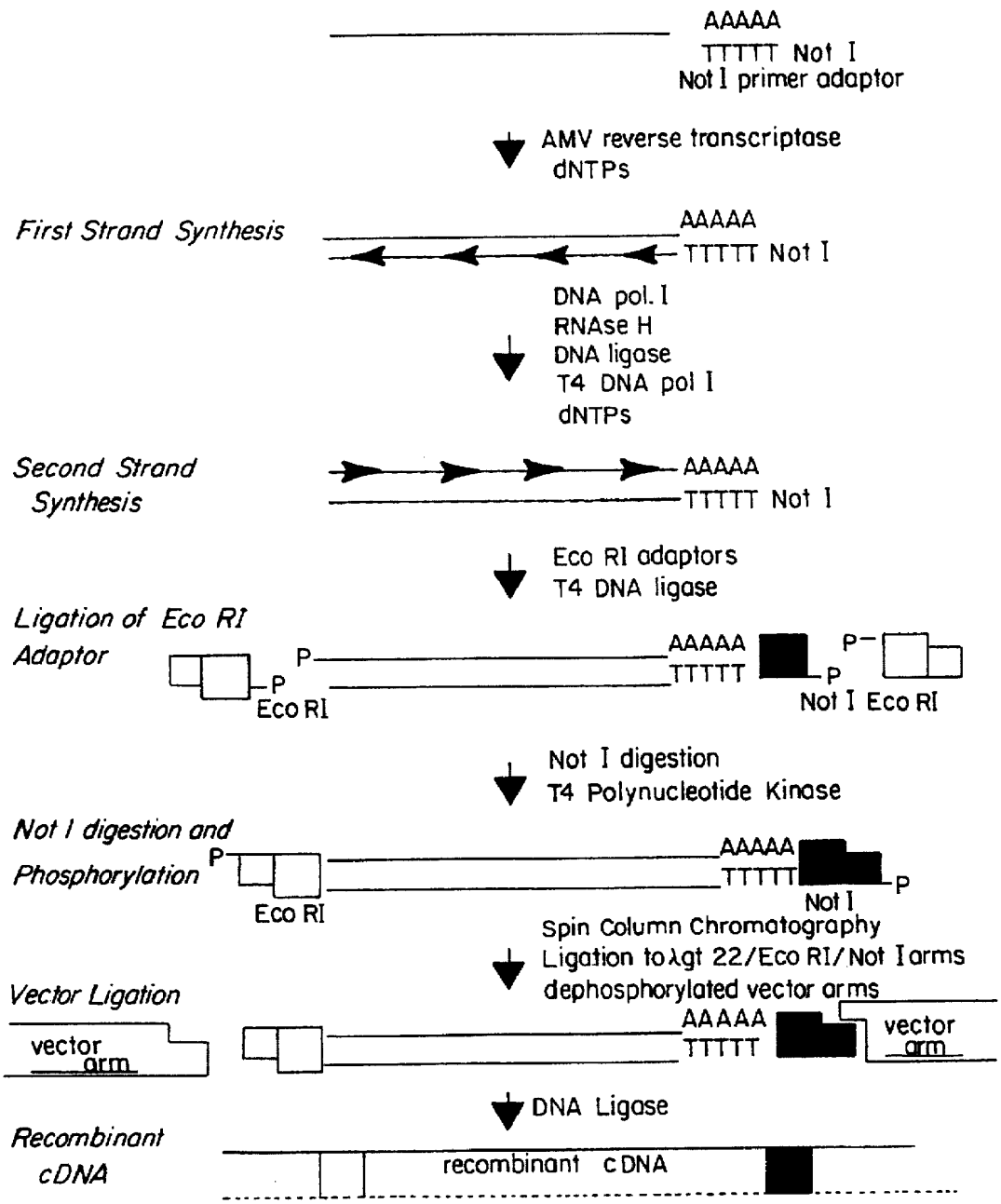

AAAAA
TTTTT Not I
Not I primer adaptor
AMV reverse transcriptase
dNTPs
First Strand Synthesis
AAAAA
TTTTT Not I
DNA pol. I
RNAse H
DNA ligase
T4 DNA pol I
dNTPs
Second Strand Synthesis
AAAAA
TTTTT Not I
Eco RI adaptors
T4 DNA ligase
Ligation of Eco RI Adaptor
P
P
Eco RI
AAAAA
TTTTT
P
P
Not I Eco RI
Not I digestion
T4 Polynucleotide Kinase
Not I digestion and Phosphorylation
P
Eco RI
AAAAA
TTTTT
Not I
P
Spin Column Chromatography
Ligation to λgt 22/Eco RI/Not I arms
dephosphorylated vector arms
Vector Ligation
vector arm
AAAAA
TTTTT
vector arm
DNA Ligase
Recombinant cDNA
recombinant cDNA

FIG. 2

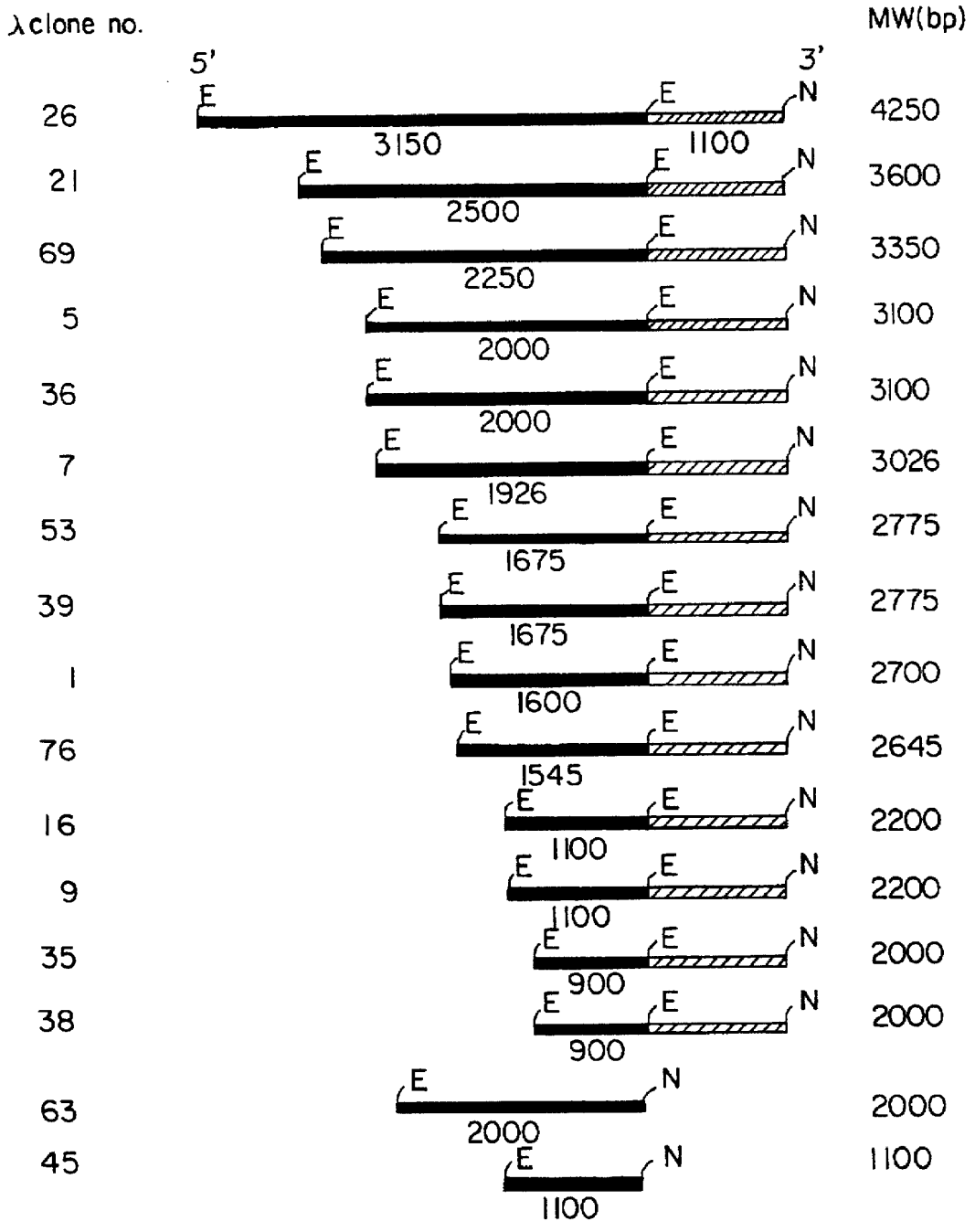

λclone no.
MW(bp)
5'
3'
26
4250
3150
1100
21
3600
2500
69
3350
2250
5
3100
2000
36
3100
2000
7
3026
1926
53
2775
1675
39
2775
1675
1
2700
1600
76
2645
1545
16
2200
1100
9
2200
1100
35
2000
900
38
2000
900
63
2000
2000
45
1100
1100
E
N

FIG. 3

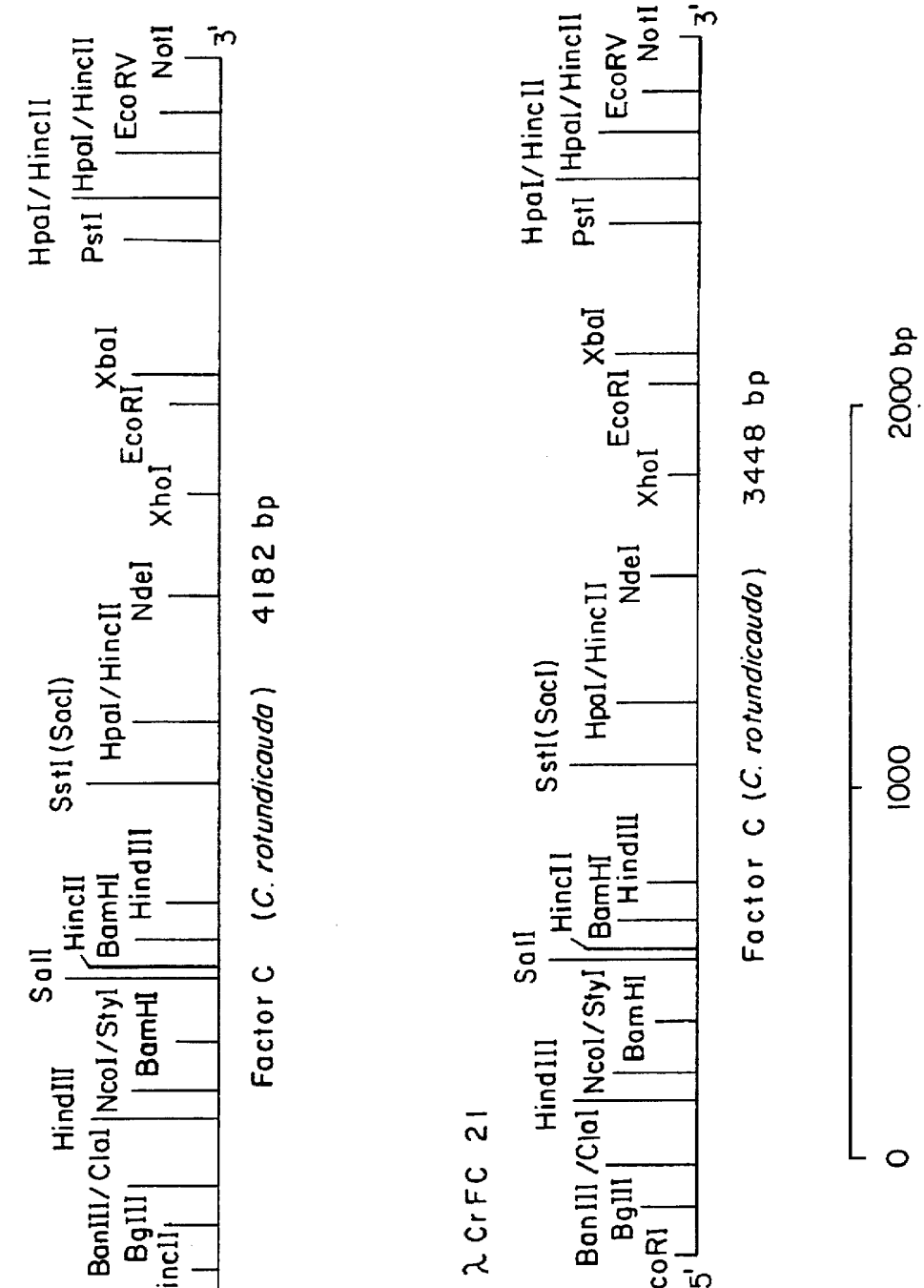

λ CrFC26
EcoRI
HincII
HincII
BglII
BanIII/ClaI
HindIII
NcoI/StyI
BamHI
SalI
HincII
BamHI
HindIII
SstI(SacI)
HpaI/HincII
NdeI
XhoI
EcoRI
XbaI
HpaI/HincII
PstI
HpaI/HincII
EcoRV
NotI
5'
3'
Factor C (C. rotundicauda) 4182 bp
λ CrFC 21
HindIII
BanIII/ClaI
BglII
EcoRI
5'
SalI
NcoI/StyI
BamHI
HincII
BamHI
HindIII
SstI(SacI)
HpaI/HincII
NdeI
XhoI
EcoRI
XbaI
HpaI/HincII
PstI
HpaI/HincII
EcoRV
NotI
3'
Factor C (C. rotundicauda) 3448 bp
0
1000
2000 bp

FIG. 4A

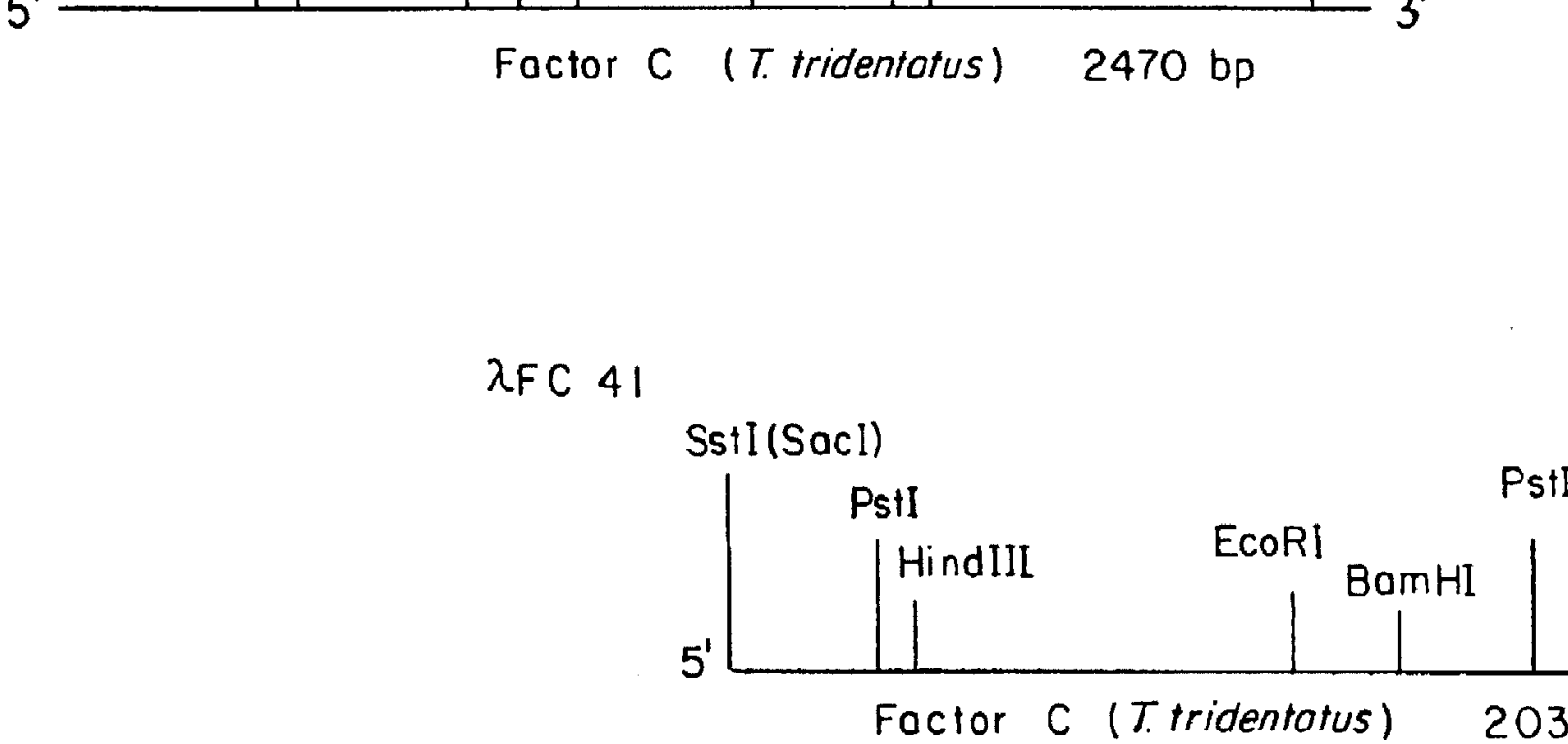


FIG. 4B

Seikagaku Corporation Exhibit 2004 - Page 226 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

FIG.5A

Seikagaku Corporation Exhibit 2004 - Page 227 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

FIG. 5B

Seikagaku Corporation Exhibit 2004 - Page 228 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

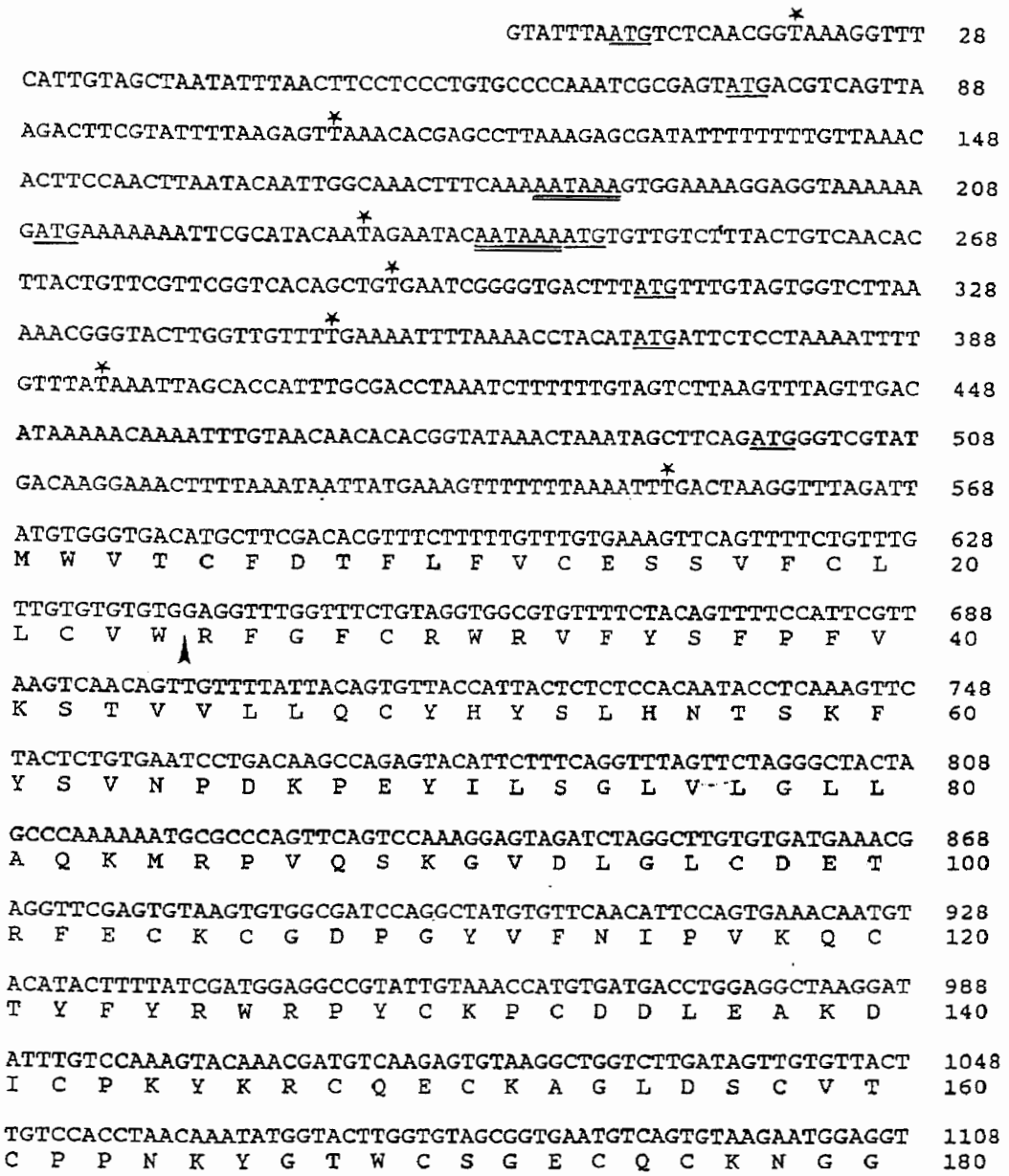

GTATTTAATGTCTCAACGGTAAAGGTTT 28
CATTGTAGCTAATATTTAACTTCCTCCCTGTGCCCCAAATCGCGAGTATGACGTCAGTTA 88
AGACTTCGTATTTTAAGAGTTAAACACGAGCCTTAAAGAGCGATATTTTTTTTGTTAAAC 148
ACTTCCAACTTAATACAATTGGCAAACTTTCAAAAATAAAGTGGAAAAGGAGGTAAAAAA 208
GATGAAAAAAATTCGCATACAATAGAATACAATAAAATGTGTTGTCTTTACTGTCAACAC 268
TTACTGTTCGTTCGGTCACAGCTGTGAATCGGGGTGACTTTATGTTTGTAGTGGTCTTAA 328
AAACGGGTACTTGGTTGTTTTGAAAATTTTAAAACCTACATATGATTCTCCTAAAATTTT 388
GTTTATAAATTAGCACCATTTGCGACCTAAATCTTTTTTGTAGTCTTAAGTTTAGTTGAC 448
ATAAAAACAAAATTTGTAACAACACACGGTATAAACTAAATAGCTTCAGATGGGTCGTAT 508
GACAAGGAAACTTTTAAATAATTATGAAAGTTTTTTTAAAATTTGACTAAGGTTTAGATT 568
ATGTGGGTGACATGCTTCGACACGTTTCTTTTTGTTTGTGAAAGTTCAGTTTTCTGTTTG 628
M W V T C F D T F L F V C E S S V F C L 20
TTGTGTGTGTGGAGGTTTGGTTTCTGTAGGTGGCGTGTTTTCTACAGTTTTCCATTCGTT 688
L C V W R F G F C R W R V F Y S F P F V 40
AAGTCAACAGTTGTTTTATTACAGTGTTACCATTACTCTCTCCACAATACCTCAAAGTTC 748
K S T V V L L Q C Y H Y S L H N T S K F 60
TACTCTGTGAATCCTGACAAGCCAGAGTACATTCTTTCAGGTTTAGTTCTAGGGCTACTA 808
Y S V N P D K P E Y I L S G L V L G L L 80
GCCCAAAAAATGCGCCCAGTTCAGTCCAAAGGAGTAGATCTAGGCTTGTGTGATGAAACG 868
A Q K M R P V Q S K G V D L G L C D E T 100
AGGTTCGAGTGTAAGTGTGGCGATCCAGGCTATGTGTTCAACATTCCAGTGAAACAATGT 928
R F E C K C G D P G Y V F N I P V K Q C 120
ACATACTTTTATCGATGGAGGCCGTATTGTAAACCATGTGATGACCTGGAGGCTAAGGAT 988
T Y F Y R W R P Y C K P C D D L E A K D 140
ATTTGTCCAAAGTACAAACGATGTCAAGAGTGTAAGGCTGGTCTTGATAGTTGTGTTACT 1048
I C P K Y K R C Q E C K A G L D S C V T 160
TGTCCACCTAACAAATATGGTACTTGGTGTAGCGGTGAATGTCAGTGTAAGAATGGAGGT 1108
C P P N K Y G T W C S G E C Q C K N G G 180

FIG. 6A

```
ATCTGTGACCAGAGGACAGGAGCTTGTGCATGTCGTGACAGATATGAAGGGGTGCACTGT  1168
I  C  D  Q  R  T  G  A  C  A  C  R  D  R  Y  E  G  V  H  C     200

GAAATTCTCAAAGGTTGTCCTCTTCTTCCATCGGATTCTCAGGTTCAGGAAGTCAGAAAT  1228
E  I  L  K  G  C  P  L  L  P  S  D  S  Q  V  Q  E  V  R  N     220

CCACCAGATAATCCCCAAACTATTGACTACAGCTGTTCACCAGGGTTCAAGCTTAAGGGT  1288
P  P  D  N  P  Q  T  I  D  Y  S  C  S  P  G  F  K  L  K  G     240

ATGGCACGAATTAGCTGTCTCCCAAATGGACAGTGGAGTAACTTTCCACCCAAATGTATT  1348
M  A  R  I  S  C  L  P  N  G  Q  W  S  N  F  P  P  K  C  I     260

CGAGAATGTGCCATGGTTTCATCTCCAGAACATGGGAAAGTGAATGCTCTTAGTGGTGAT  1408
R  E  C  A  M  V  S  S  P  E  H  G  K  V  N  A  L  S  G  D     280

ATGATAGAAGGGGCTACTTTACGGTTCTCATGTGATAGTCCCTACTACTTGATTGGTCAA  1468
M  I  E  G  A  T  L  R  F  S  C  D  S  P  Y  Y  L  I  G  Q     300

GAAACATTAACCTGTCAGGGTAATGGTCAGTGGAATGGACAGATACCACAATGTAAGAAC  1528
E  T  L  T  C  Q  G  N  G  Q  W  N  G  Q  I  P  Q  C  K  N     320

TTAGTCTTCTGTCCTGACCTGGATCCTGTAAACCATGCTGAACACAAGGTTAAAATTGGT  1588
L  V  F  C  P  D  L  D  P  V  N  H  A  E  H  K  V  K  I  G     340

GTGGAACAAAAATATGGTCAGTTTCCTCAAGGCACTGAAGTGACCTATACGTGTTCGGGT  1648
V  E  Q  K  Y  G  Q  F  P  Q  G  T  E  V  T  Y  T  C  S  G     360

AACTACTTCTTGATGGGTTTTGACACCTTAAAATGTAACCCTGATGGGTCTTGGTCAGGA  1708
N  Y  F  L  M  G  F  D  T  L  K  C  N  P  D  G  S  W  S  G     380

TCACAGCCATCCTGTGTTAAAGTGGCAGACAGAGAGGTCGACTGTGACAGTAAAGCTGTA  1768
S  Q  P  S  C  V  K  V  A  D  R  E  V  D  C  D  S  K  A  V     400

GACTTCTTGGATGATGTTGGTGAACCTGTCAGGATCCACTGTCCTGCTGGCTGTTCTTTG  1828
D  F  L  D  D  V  G  E  P  V  R  I  H  C  P  A  G  C  S  L     420

ACAGCTGGTACTGTGTGGGGTACAGCCATATACCATGAACTTTCCTCAGTGTGTCGTGCA  1888
T  A  G  T  V  W  G  T  A  I  Y  H  E  L  S  S  V  C  R  A     440

GCCATCCATGCTGGCAAGCTTCCAAACTCTGGAGGAGCGGTGCATGTTGTGAACAATGGC  1948
A  I  H  A  G  K  L  P  N  S  G  G  A  V  H  V  V  N  N  G     460

CCCTACTCGGACTTTCTGGGTAGTGACCTGAATGGGATAAAATCCGAAGAGTTGAAGTCT  2008
P  Y  S  D  F  L  G  S  D  L  N  G  I  K  S  E  E  L  K  S     480

CTTGCCCGGAGTTTCCGATTCGATTATGTCAGTTCCTCCACAGCAGGTAAATCAGGATGT  2068
L  A  R  S  F  R  F  D  Y  V  S  S  S  T  A  G  K  S  G  C     500

CCTGATGGATGGTTTGAGGTAGACGAGAACTGTGTGTACGTTACATCAAAACAGAGAGCC  2128
P  D  G  W  F  E  V  D  E  N  C  V  Y  V  T  S  K  Q  R  A     520
```

FIG. 6B

```
TGGGAAAGAGCTCAAGGTGTGTGTACCAATATGGCTGCTCGTCTTGCTGTGCTGGACAAA  2188
W  E  R  A  Q  G  V  C  T  N  M  A  A  R  L  A  V  L  D  K    540

GATGTAATTCCAAATTCATTGACTGAGACTCTACGAGGGAAAGGGTTAACAACCACGTGG  2248
D  V  I  P  N  S  L  T  E  T  L  R  G  K  G  L  T  T  T  W    560

ATAGGATTGCACAGACTAGATGCTGAGAAGCCCTTTATTTGGGAGTTAATGGATCGTAGT  2308
I  G  L  H  R  L  D  A  E  K  P  F  I  W  E  L  M  D  R  S    580

AATGTGGTTCTGAATGATAACCTAACATTCTGGGCCTCTGGCGAACCTGGAAATGAAACT  2368
N  V  V  L  N  D  N  L  T  F  W  A  S  G  E  P  G  N  E  T    600

AACTGTGTATATATGGACATCCAAGATCAGTTGCAGTCTGTGTGGAAAACCAAGTCATGT  2428
N  C  V  Y  M  D  I  Q  D  Q  L  Q  S  V  W  K  T  K  S  C    620

TTTCAGCCCTCAAGTTTTGCTTGCATGATGGATCTGTCAGACAGAAATAAAGCCAAATGC  2488
F  Q  P  S  S  F  A  C  M  M  D  L  S  D  R  N  K  A  K  C    640

GATGATCCTGGATCACTGGAAAATGGACACGCCACACTTCATGGACAAAGTATTGATGGG  2548
D  D  P  G  S  L  E  N  G  H  A  T  L  H  G  Q  S  I  D  G    660

TTCTATGCTGGTTCTTCTATAAGGTACAGCTGTGAGGTTCTCCACTACCTCAGTGGAACT  2608
F  Y  A  G  S  S  I  R  Y  S  C  E  V  L  H  Y  L  S  G  T    680

GAAACCGTAACTTGTACAACAAATGGCACATGGAGTGCTCCTAAACCTCGATGTATCAAA  2668
E  T  V  T  C  T  T  N  G  T  W  S  A  P  K  P  R  C  I  K    700

GTCATCACCTGCCAAAACCCCCCTGTACCATCATATGGTTCTGTGGAAATCAAACCCCCA  2728
V  I  T  C  Q  N  P  P  V  P  S  Y  G  S  V  E  I  K  P  P    720

AGTCGGACAAACTCGATAAGTCGTGTTGGGTCACCTTTCTTGAGGTTGCCACGGTTACCC  2788
S  R  T  N  S  I  S  R  V  G  S  P  F  L  R  L  P  R  L  P    740

CTCCCATTAGCCAGAGCAGCCAAACCTCCTCCAAAACCTAGATCCTCACAACCCTCTACT  2848
L  P  L  A  R  A  A  K  P  P  P  K  P  R  S  S  Q  P  S  T    760

GTGGACTTGGCTTCTAAAGTTAAACTACCTGAAGGTCATTACCGGGTAGGGTCTCGAGCC  2908
V  D  L  A  S  K  V  K  L  P  E  G  H  Y  R  V  G  S  R  A    780

ATTTACACGTGCGAGTCGAGATACTACGAACTACTTGGATCTCAAGGCAGAAGATGTGAC  2968
I  Y  T  C  E  S  R  Y  Y  E  L  L  G  S  Q  G  R  R  C  D    800

TCTAATGGAAACTGGAGTGGTCGGCCAGCGAGCTGTATTCCAGTTTGTGGACGGTCAGAC  3028
S  N  G  N  W  S  G  R  P  A  S  C  I  P  V  C  G  R  S  D    820

TCTCCTCGTTCTCCTTTTATCTGGAATGGGAATTCTACAGAAATAGGTCAGTGGCCGTGG  3088
S  P  R  S  P  F  I  W  N  G  N  S  T  E  I  G  Q  W  P  W    840

CAGGCAGGAATCTCTAGATGGCTTGCAGACCACAATATGTGGTTTCTCCAGTGTGGAGGA  3148
Q  A  G  I  S  R  W  L  A  D  H  N  M  W  F  L  Q  C  G  G    860

TCTCTATTGAATGAGAAATGGATCGTCACTGCTGCCCACTGTGTCACCTACTCTGCTACT  3208
S  L  L  N  E  K  W  I  V  T  A  A  H  C  V  T  Y  S  A  T    880
```

FIG. 6C

```
GCTGAGATTATTGACCCCAATCAGTTTAAAATGTATCTGGGCAAGTACTACCGTGATGAC  3268
A  E  I  I  D  P  N  Q  F  K  M  Y  L  G  K  Y  Y  R  D  D     900

AGTAGAGACGATGACTATGTACAAGTAAGAGAGGCTCTTGAGATCCACGTGAATCCTAAC  3328
S  R  D  D  D  Y  V  Q  V  R  E  A  L  E  I  H  V  N  P  N     920

TACGACCCCGGCAATCTCAACTTTGACATAGCCCTAATTCAACTGAAAACTCCTGTTACT  3388
Y  D  P  G  N  L  N  F  D  I  A  L  I  Q  L  K  T  P  V  T     940
                          ✪

TTGACAACACGAGTCCAACCAATCTGTCTGCCTACTGACATCACAACAAGAGAACACTTG  3448
L  T  T  R  V  Q  P  I  C  L  P  T  D  I  T  T  R  E  H  L     960

AAGGAGGGAACATTAGCAGTGGTGACAGGTTGGGGTTTGAATGAAAACAACACCTATTCA  3508
K  E  G  T  L  A  V  V  T  G  W  G  L  N  E  N  N  T  Y  S     980
                                   ◆

GAGACGATTCAACAAGCTGTGCTACCTGTTGTTGCAGCCAGCACCTGTGAAGAGGGGTAC  3568
E  T  I  Q  Q  A  V  L  P  V  V  A  A  S  T  C  E  E  G  Y     1000

AAGGAAGCAGACTTACCACTGACAGTAACAGAGAACATGTTCTGTGCAGGTTACAAGAAG  3628
K  E  A  D  L  P  L  T  V  T  E  N  M  F  C  A  G  Y  K  K     1020

GGACGTTATGATGCCTGCAGTGGGGACAGTGGAGGACCTTTAGTGTTTGCTGATGATTCC  3688
G  R  Y  D  A  C  S  G  D  S  G  G  P  L  V  F  A  D  D  S     1040
                          ✪

CGTACCGAAAGGCGGTGGGTCTTGGAAGGGATTGTCAGCTGGGGCAGTCCCAGTGGATGT  3748
R  T  E  R  R  W  V  L  E  G  I  V  S  W  G  S  P  S  G  C     1060

GGCAAGGCGAACCAGTACGGGGGCTTCACTAAAGTTAACGTTTTCCTGTCATGGATTAGG  3808
G  K  A  N  Q  Y  G  G  F  T  K  V  N  V  F  L  S  W  I  R     1080

CAGTTCATTTGAAACTGATCTAAATATTTTAAGCATGGTTATAAACGTCTTGTTCCTATT  3868
Q  F  I  <        1083

ATTGCTTTACTGGTTTAACCCATAAGAAGGTTAACGGGGTAAGGCACAAGGATCATTGTT  3928

TCTGTTTGTTTTTACAAATGGTTCTTTTAGTCAGTGAATGAGAATAGTATCCATTGGAGA  3988

CTGTTACCTTTTATTCTACCTTTTTATATTACTATGCAAGTATTTGGGATATCTTCTACA  4048

CATGAAAATTCTGTCATTTTACCATAAATTTGGTTTCTGGTGTGTGTGTTAAGTCCACCA  4108

CTAGAGAACGATGTAATTTTCAATAGTACATGAAATAAATATAGAACAAATCTATTATAA  4168

AAAAAAAAAAAAAA                                                 4182
```

FIG. 6D

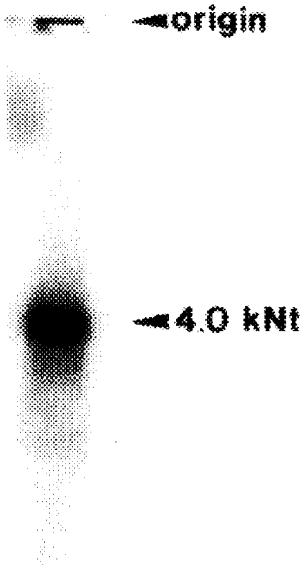


FIG.7A

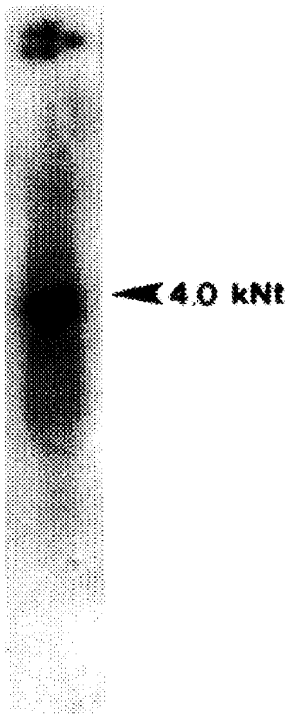


FIG.7B

```
                                                    GTGAAGGTAACTTAAGT  17

ATGGTCTTAGCGTCGTTTTTGGTGTCTGGTTTAGTTCTAGGGCTACTAGCCCAAAAAATG  77
M  V  L  A  S  F  L  V  S  G  L  V  L  G  L  L  A  Q  K  M    20

CGCCCAGTTCAGTCCAAAGGAGTAGATCTAGGCTTGTGTGATGAAACGAGGTTCGAGTGT  137
R  P  V  Q  S  K  G  V  D  L  G  L  C  D  E  T  R  F  E  C    40

AAGTGTGGCGATCCAGGCTATGTGTTCAACATTCCAGTGAAACAATGTACATACTTTTAT  197
K  C  G  D  P  G  Y  V  F  N  I  P  V  K  Q  C  T  Y  F  Y    60

CGATGGAGGCCGTATTGTAAACCATGTGATGACCTGGAGGCTAAGGATATTTGTCCAAAG  257
R  W  R  P  Y  C  K  P  C  D  D  L  E  A  K  D  I  C  P  K    80

TACAAACGATGTCAAGAGTGTAAGGCTGGTCTTGATAGTTGTGTTACTTGTCCACCTAAC  317
Y  K  R  C  Q  E  C  K  A  G  L  D  S  C  V  T  C  P  P  N    100

AAATATGGTACTTGGTGTAGCGGTGAATGTCAGTGTAAGAATGGAGGTATCTGTGACCAG  377
K  Y  G  T  W  C  S  G  E  C  Q  C  K  N  G  G  I  C  D  Q    120

AGGACAGGAGCTTGTGCATGTCGTGACAGATATGAAGGGGTGCACTGTGAAATTCTCAAA  437
R  T  G  A  C  A  C  R  D  R  Y  E  G  V  H  C  E  I  L  K    140

GGTTGTCCTCTTCTTCCATCGGATTCTCAGGTTCAGGAAGTCAGAAATCCACCAGATAAT  497
G  C  P  L  L  P  S  D  S  Q  V  Q  E  V  R  N  P  P  D  N    160

CCCCAAACTATTGACTACAGCTGTTCACCAGGGTTCAAGCTTAAGGGTATGGCACGAATT  557
P  Q  T  I  D  Y  S  C  S  P  G  F  K  L  K  G  M  A  R  I    180

AGCTGTCTCCCAAATGGACAGTGGAGTAACTTTCCACCCAAATGTATTCGAGAATGTGCC  617
S  C  L  P  N  G  Q  W  S  N  F  P  P  K  C  I  R  E  C  A    200

ATGGTTTCATCTCCAGAACATGGGAAAGTGAATGCTCTTAGTGGTGATATGATAGAAGGG  677
M  V  S  S  P  E  H  G  K  V  N  A  L  S  G  D  M  I  E  G    220

GCTACTTTACGGTTCTCATGTGATAGTCCCTACTACTTGATTGGTCAAGAAACATTAACC  737
A  T  L  R  F  S  C  D  S  P  Y  Y  L  I  G  Q  E  T  L  T    240

TGTCAGGGTAATGGTCAGTGGAATGGACAGATACCACAATGTAAGAACTTGGTCTTCTGT  797
C  Q  G  N  G  Q  W  N  G  Q  I  P  Q  C  K  N  L  V  F  C    260

CCTGACCTGGATCCTGTAAACCATGCTGAACACAAGGTTAAAATTGGTGTGGAACAAAAA  857
P  D  L  D  P  V  N  H  A  E  H  K  V  K  I  G  V  E  Q  K    280

TATGGTCAGTTTCCTCAAGGCACTGAAGTGACCTATACGTGTTCGGGTAACTACTTCTTG  917
Y  G  Q  F  P  Q  G  T  E  V  T  Y  T  C  S  G  N  Y  F  L    300

ATGGGTTTTGACACCTTAAAATGTAACCCTGATGGGTCTTGGTCAGGATCACAGCCATCC  977
M  G  F  D  T  L  K  C  N  P  D  G  S  W  S  G  S  Q  P  S    320

TGTGTTAAAGTGGCAGACAGAGAGGTCGACTGTGACAGTAAAGCTGTAGACTTCTTGGAT  1037
C  V  K  V  A  D  R  E  V  D  C  D  S  K  A  V  D  F  L  D    340
```

FIG. 8A

```
GATGTTGGTGAACCTGTCAGGATCCACTGTCCTGCTGGCTGTTCTTTGACAGCTGGTACT  1097
D  V  G  E  P  V  R  I  H  C  P  A  G  C  S  L  T  A  G  T     360

GTGTGGGGTACAGCCATATACCATGAACTTTCCTCAGTGTGTCGTGCAGCCATCCATGCT  1157
V  W  G  T  A  I  Y  H  E  L  S  S  V  C  R  A  A  I  H  A     380

GGCAAGCTTCCAAACTCTGGAGGAGCGGTGCATGTTGTGAACAATGGCCCCTACTCGGAC  1217
G  K  L  P  N  S  G  G  A  V  H  V  V  N  N  G  P  Y  S  D     400

TTTCTGGGTAGTGACCTGAATGGGATAAAATCGGAAGAGTTGAAGTCTCTTGCCCGGAGT  1277
F  L  G  S  D  L  N  G  I  K  S  E  E  L  K  S  L  A  R  S     420

TTCCGATTCGATTATGTCCGTTCCTCCACAGCAGGTAAATCAGGATGTCCTGATGGATGG  1337
F  R  F  D  Y  V  R  S  S  T  A  G  K  S  G  C  P  D  G  W     440

TTTGAGGTAGACGAGAACTGTGTGTACGTTACATCAAAACAGAGAGCCTGGGAAAGAGCT  1397
F  E  V  D  E  N  C  V  Y  V  T  S  K  Q  R  A  W  E  R  A     460

CAAGGTGTGTGTACCAATATGGCTGCTCGTCTTGCTGTGCTGGACAAAGATGTAATTCCA  1457
Q  G  V  C  T  N  M  A  A  R  L  A  V  L  D  K  D  V  I  P     480

AATTCGTTGACTGAGACTCTACGAGGGAAAGGGTTAACAACCACGTGGATAGGATTGCAC  1517
N  S  L  T  E  T  L  R  G  K  G  L  T  T  T  W  I  G  L  H     500

AGACTAGATGCTGAGAAGCCCTTTATTTGGGAGTTAATGGATCGTAGTAATGTGGTTCTG  1577
R  L  D  A  E  K  P  F  I  W  E  L  M  D  R  S  N  V  V  L     520

AATGATAACCTAACATTCTGGGCCTCTGGCGAACCTGGAAATGAAACTAACTGTGTATAT  1637
N  D  N  L  T  F  W  A  S  G  E  P  G  N  E  T  N  C  V  Y     540
      ◆                             ◆

ATGGACATCCAAGATCAGTTGCAGTCTGTGTGGAAAACCAAGTCATGTTTTCAGCCCTCA  1697
M  D  I  Q  D  Q  L  Q  S  V  W  K  T  K  S  C  F  Q  P  S     560

AGTTTTGCTTGCATGATGGATCTGTCAGACAGAAATAAAGCCAAATGCGATGATCCTGGA  1757
S  F  A  C  M  M  D  L  S  D  R  N  K  A  K  C  D  D  P  G     580

TCACTGGAAAATGGACACGCCACACTTCATGGACAAAGTATTGATGGGTTCTATGCTGGT  1817
S  L  E  N  G  H  A  T  L  H  G  Q  S  I  D  G  F  Y  A  G     600

TCTTCTATAAGGTACAGCTGTGAGGTTCTCCACTACCTCAGTGGAACTGAAACCGTAACT  1877
S  S  I  R  Y  S  C  E  V  L  H  Y  L  S  G  T  E  T  V  T     620

TGTACAACAAATGGCACATGGAGTGCTCCTAAACCTCGATGTATCAAAGTCATCACCTGC  1937
C  T  T  N  G  T  W  S  A  P  K  P  R  C  I  K  V  I  T  C     640
         ◆

CAAAACCCCCCCTGTACCATCATATGGTTCTGTGGAAATCAAACCCCCAAGTCGGACAAAC  1997
Q  N  P  P  V  P  S  Y  G  S  V  E  I  K  P  P  S  R  T  N     660

TCGATAAGTCGTGTTGGGTCACCTTTCTTGAGGTTGCCACGGTTACCCCTCCCATTAGCT  2057
S  I  S  R  V  G  S  P  F  L  R  L  P  R  L  P  L  P  L  A     680

AGAGCAGCCAAACCTCCTCCAAAACCTAGATCCTCACAACCCTCTACTGTGGACTTGGCT  2117
R  A  A  K  P  P  P  K  P  R  S  S  Q  P  S  T  V  D  L  A     700
                          ⬠

TCTAAAGTTAAACTACCTGAAGGTCATTACCGGGTAGGGTCTCGAGCCATCTACACGTGC  2177
S  K  V  K  L  P  E  G  H  Y  R  V  G  S  R  A  I  Y  T  C     720

GAGTCGAGATACTACGAACTACTTGGATCTCAAGGCAGAAGATGTGACTCTAATGGAAAC  2237
E  S  R  Y  Y  E  L  L  G  S  Q  G  R  R  C  D  S  N  G  N     740
                                                      ◆
```

FIG. 8B

```
TGGAGTGGTCGGCCAGCGAGCTGTATTCCAGTTTGTGGACGGTCAGACTCTCCTCGTTCT  2297
W  S  G  R  P  A  S  C  I  P  V  C  G  R  S  D  S  P  R  S     760

CCTTTTATCTGGAATGGGAATTCTACAGAAATAGGTCAGTGGCCGTGGCAGGCAGGAATC  2357
P  F  I  W  N  G  N  S  T  E  I  G  Q  W  P  W  Q  A  G  I     780

TCTAGATGGCTTGCAGACCACAATATGTGGTTTCTCCAGTGTGGAGGATCTCTATTGAAT  2417
S  R  W  L  A  D  H  N  M  W  F  L  Q  C  G  G  S  L  L  N     800

GAGAAATGGATCGTCACTGCTGCCCACTGTGTCACCTACTCTGCTACTGCTGAGATTATT  2477
E  K  W  I  V  T  A  A  H  C  V  T  Y  S  A  T  A  E  I  I     820

GACCCCAATCAGTTTAAAATGTATCTGGGCAAGTACTACCGTGATGACAGTAGAGACGAT  2537
D  P  N  Q  F  K  M  Y  L  G  K  Y  Y  R  D  D  S  R  D  D     840

GACTATGTACAAGTAAGAGAGGCTCTTGAGATCCACGTGAATCCTAACTACGACCCCGGC  2597
D  Y  V  Q  V  R  E  A  L  E  I  H  V  N  P  N  Y  D  P  G     860

AATCTCAACTTTGACATAGCCCTAATTCAACTGAAAACTCCTGTTACTTTGACAACACGA  2657
N  L  N  F  D  I  A  L  I  Q  L  K  T  P  V  T  L  T  T  R     880

GTCCAACCAATCTGTCTGCCTACTGACATCACAACAAGAGAACACTTGAAGGAGGGAACA  2717
V  Q  P  I  C  L  P  T  D  I  T  T  R  E  H  L  K  E  G  T     900

TTAGCAGTGGTGACAGGTTGGGGTTTGAATGAAAACAACACCTATTCAGAGACGATTCAA  2777
L  A  V  V  T  G  W  G  L  N  E  N  N  T  Y  S  E  T  I  Q     920

CAAGCTGTGCTACCTGTTGTTGCAGCCAGCACCTGTGAAGAGGGGTACAAGGAAGCAGAC  2837
Q  A  V  L  P  V  V  A  A  S  T  C  E  E  G  Y  K  E  A  D     940

TTACCACTGACAGTAACAGAGAACATGTTCTGTGCAGGTTACAAGAAGGGACGTTATGAT  2897
L  P  L  T  V  T  E  N  M  F  C  A  G  Y  K  K  G  R  Y  D     960

GCCTGCAGTGGGGACAGTGGAGGACCTTTAGTGTTTGCTGATGATTCCCGTACCGAAAGG  2957
A  C  S  G  D  S  G  G  P  L  V  F  A  D  D  S  R  T  E  R     980

CGGTGGGTCTTGGAAGGGATTGTCAGCTGGGGCAGTCCCAGTGGATGTGGCAAGGCGAAC  3017
R  W  V  L  E  G  I  V  S  W  G  S  P  S  G  C  G  K  A  N    1000

CAGTACGGGGGCTTCACTAAAGTTAACGTTTTCCTGTCATGGATTAGGCAGTTCATTTGA  3077
Q  Y  G  G  F  T  K  V  N  V  F  L  S  W  I  R  Q  F  I  <    1019

AACTGATCTAAATATTTTAAGCATGGTTATAAACGTCTTGTTTCCTATTATTGCTTTACT  3137

AGTTTAACCCATAAGAAGGTTAACTGGGTAAGGCACAAGGATCATTGTTTCTGTTTGTTT  3197

TTACAAATGGTTATTTTAGTCAGTGAATGAGAATAGTATCCATTGAAGACTGTTACCTTT  3257

TATTCTACCTTTTTATATTACTATGTAAGTATTTGGGATATCTTCTACACATGAAAATTC  3317

TGTCATTTTACCATAAATTTGGTTTCTGGTGTGTGCTAAGTCCACCAGTAGAGAACGATG  3377

TAATTTTCACTAGCACATGAAATAAATATAGAACAAATCTATTATAAACTACCTTAAAAA  3437

AAAAAAAAAAA  3448
```

FIG. 8C

```
pCrFC21     1                                          GTGAAGGTAACTTAAGTATGGTCTTAGCGTCGTTTTT
pFC53       1   CACGTTTAACGCGAACGTGGAAGAACTCTGAAGGTAACTTAAGTATGGTCTTAGCGTCGTTTTT
pCrFC26   717   ACCATTACTCTCTCCACAATACCTCAAAGTTCTACTCTGTGAATCCTGACAAGCCAGAGTACAT
```

Fig. 9

Seikagaku Corporation Exhibit 2004 - Page 237 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

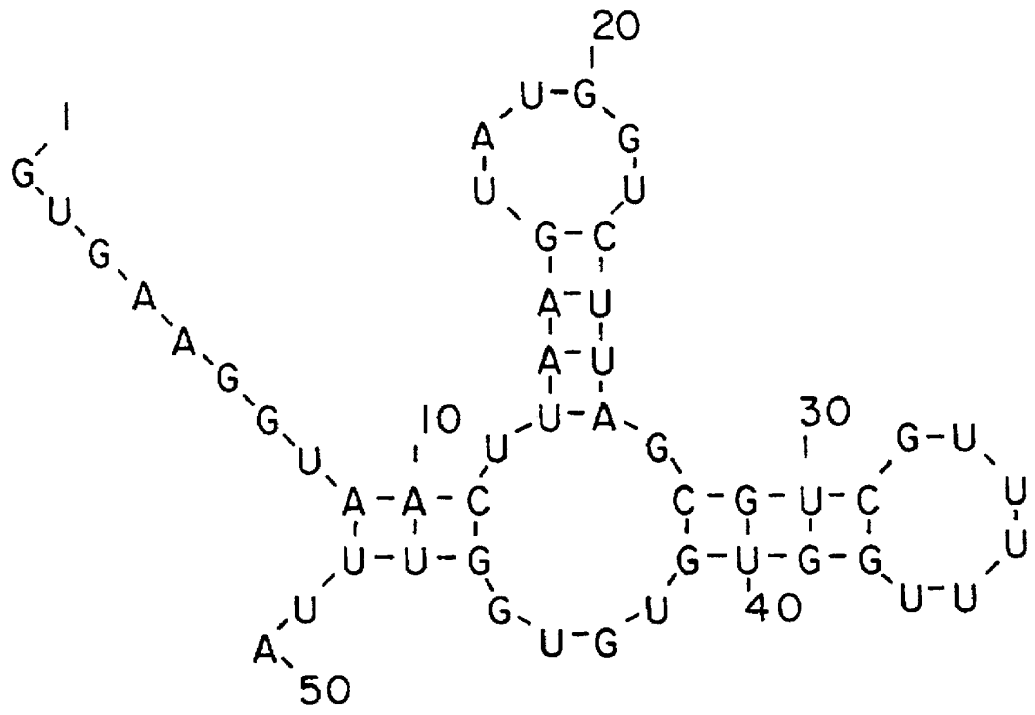


FIG. 10A

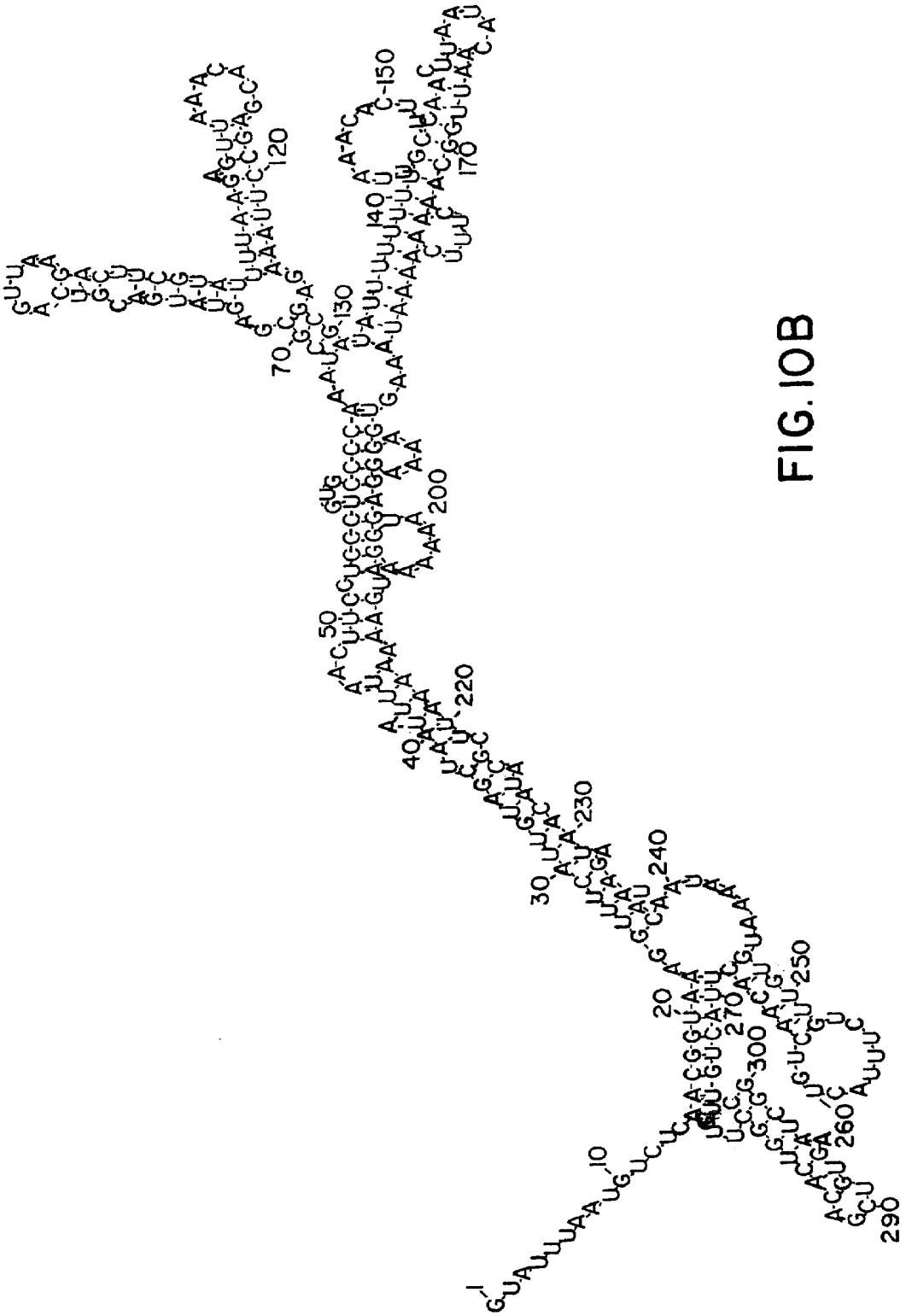


FIG. 10B

FIG. 10

Seikagaku Corporation Exhibit 2004 - Page 240 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

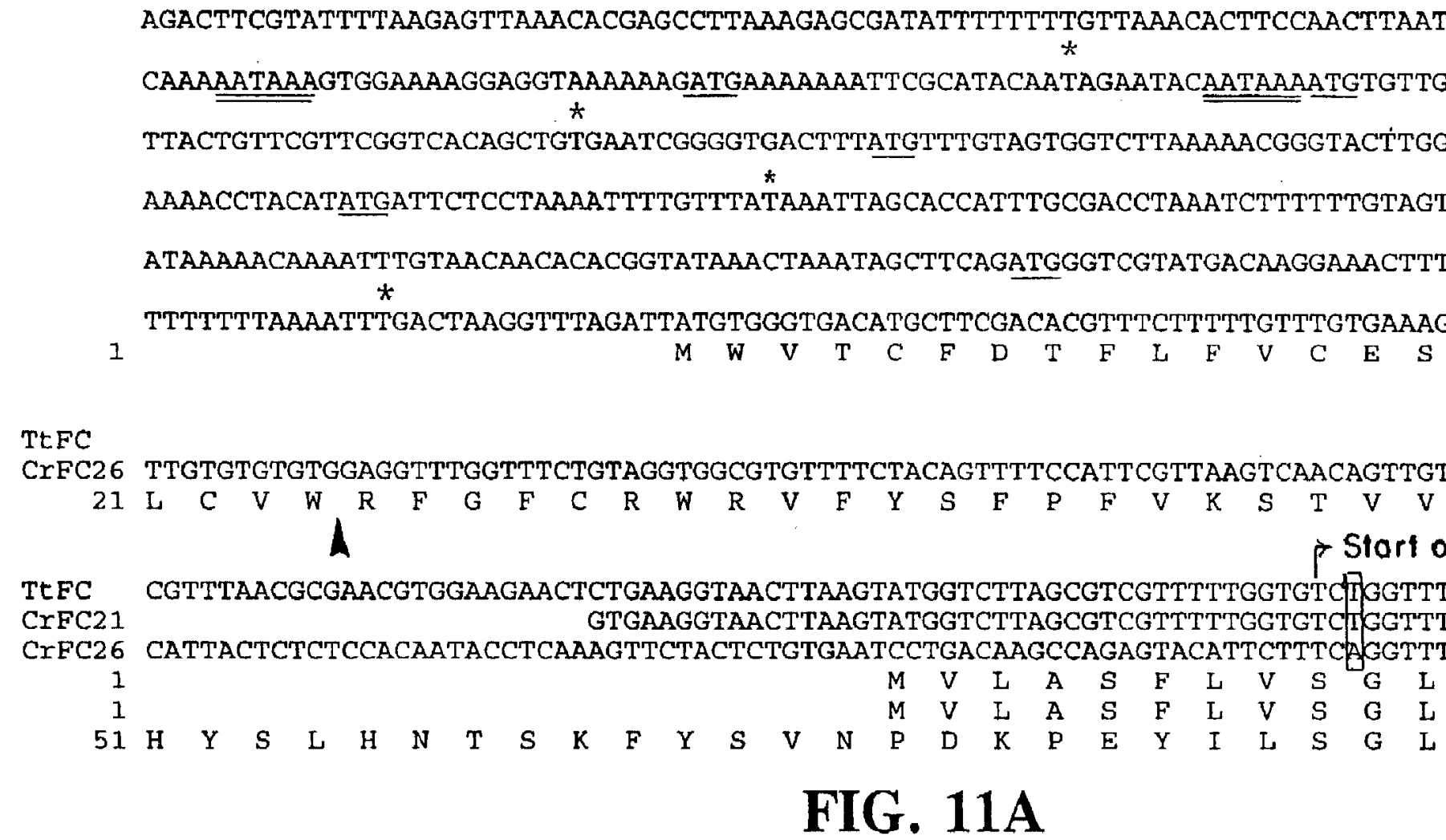


FIG. 11A

Seikagaku Corporation Exhibit 2004 - Page 241 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

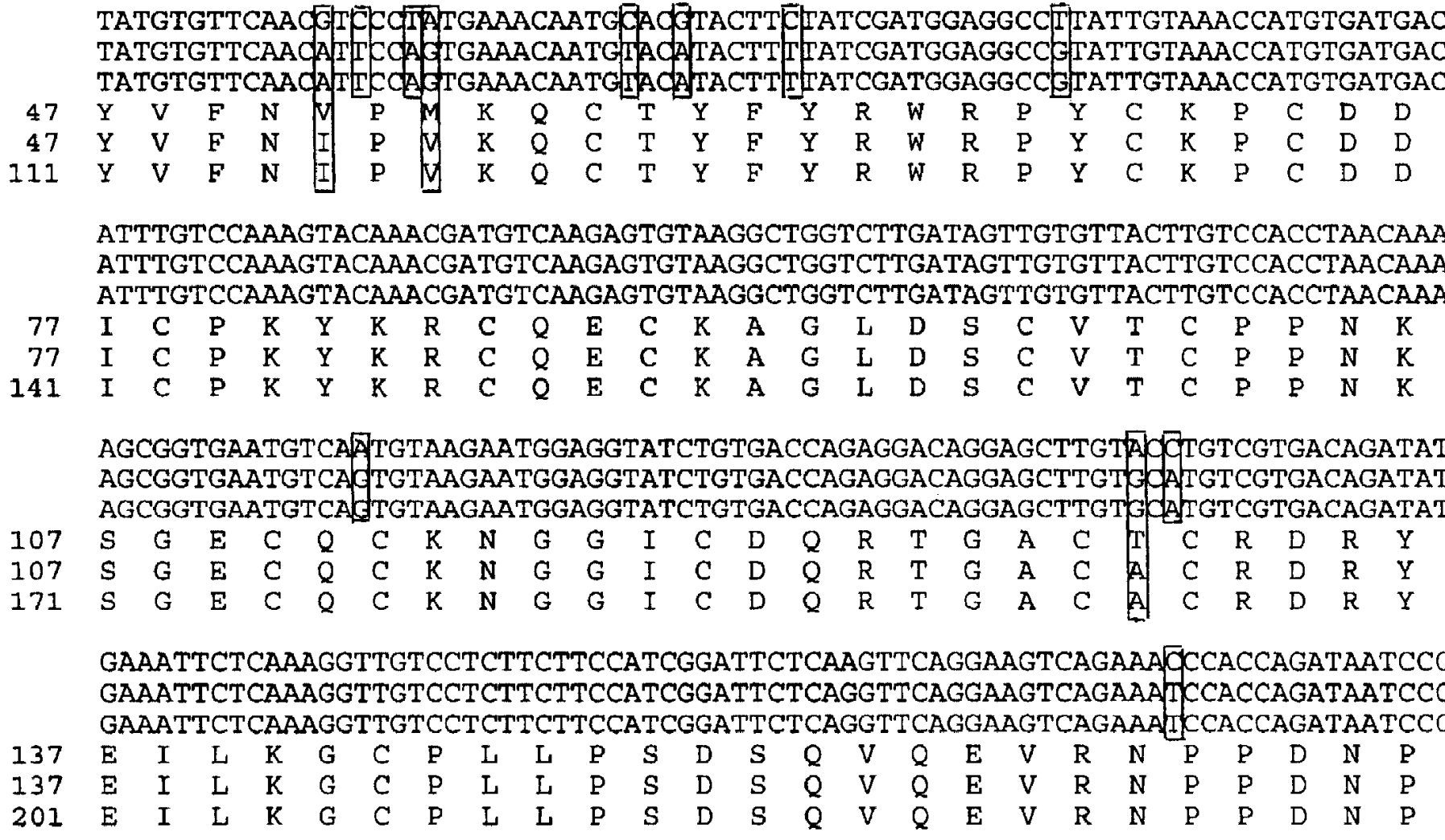


FIG. 11B

Seikagaku Corporation Exhibit 2004 - Page 242 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

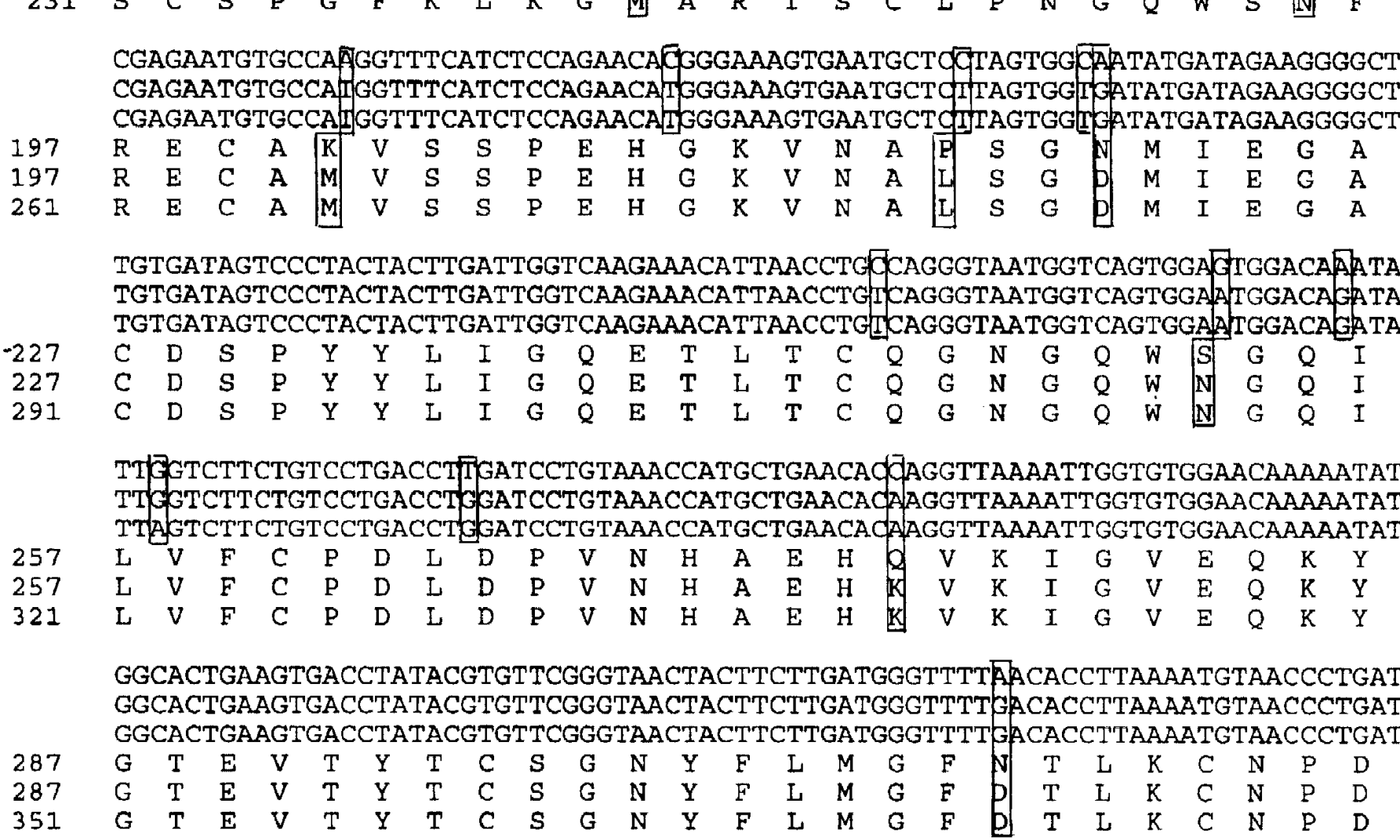


FIG. 11C

Seikagaku Corporation Exhibit 2004 - Page 243 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

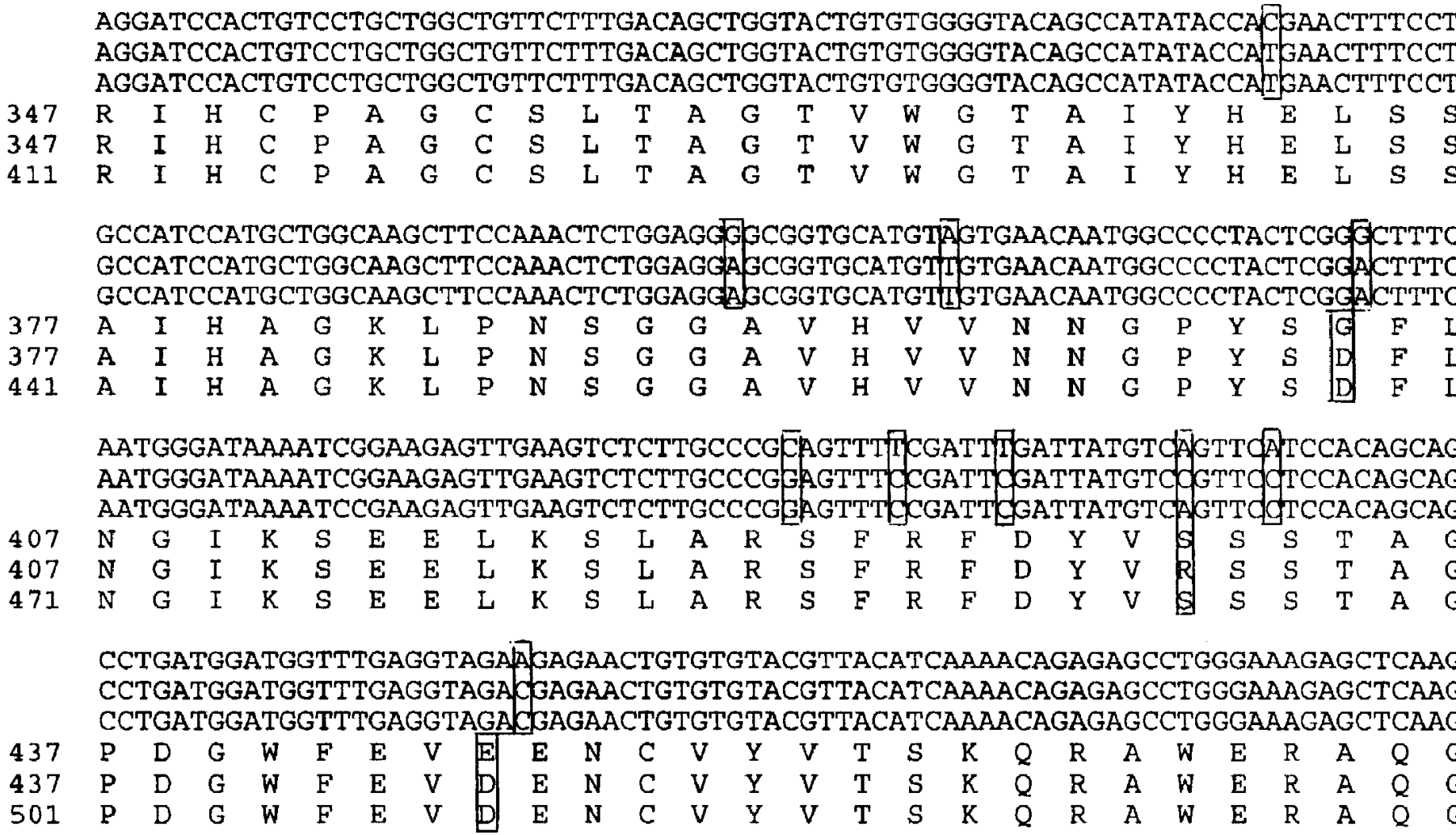


FIG. 11D

Seikagaku Corporation Exhibit 2004 - Page 244 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

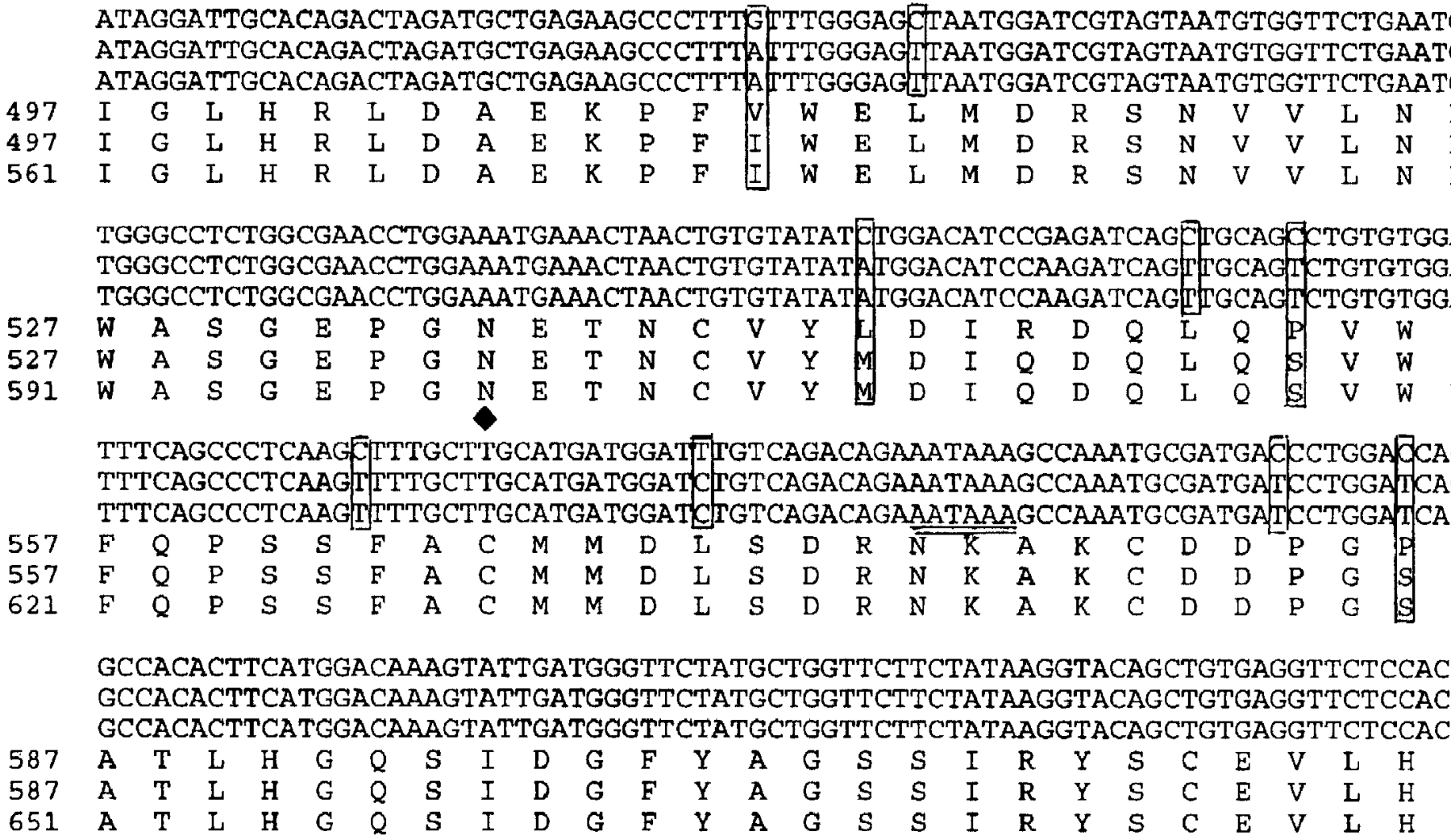


FIG. 11E

Seikagaku Corporation Exhibit 2004 - Page 245 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

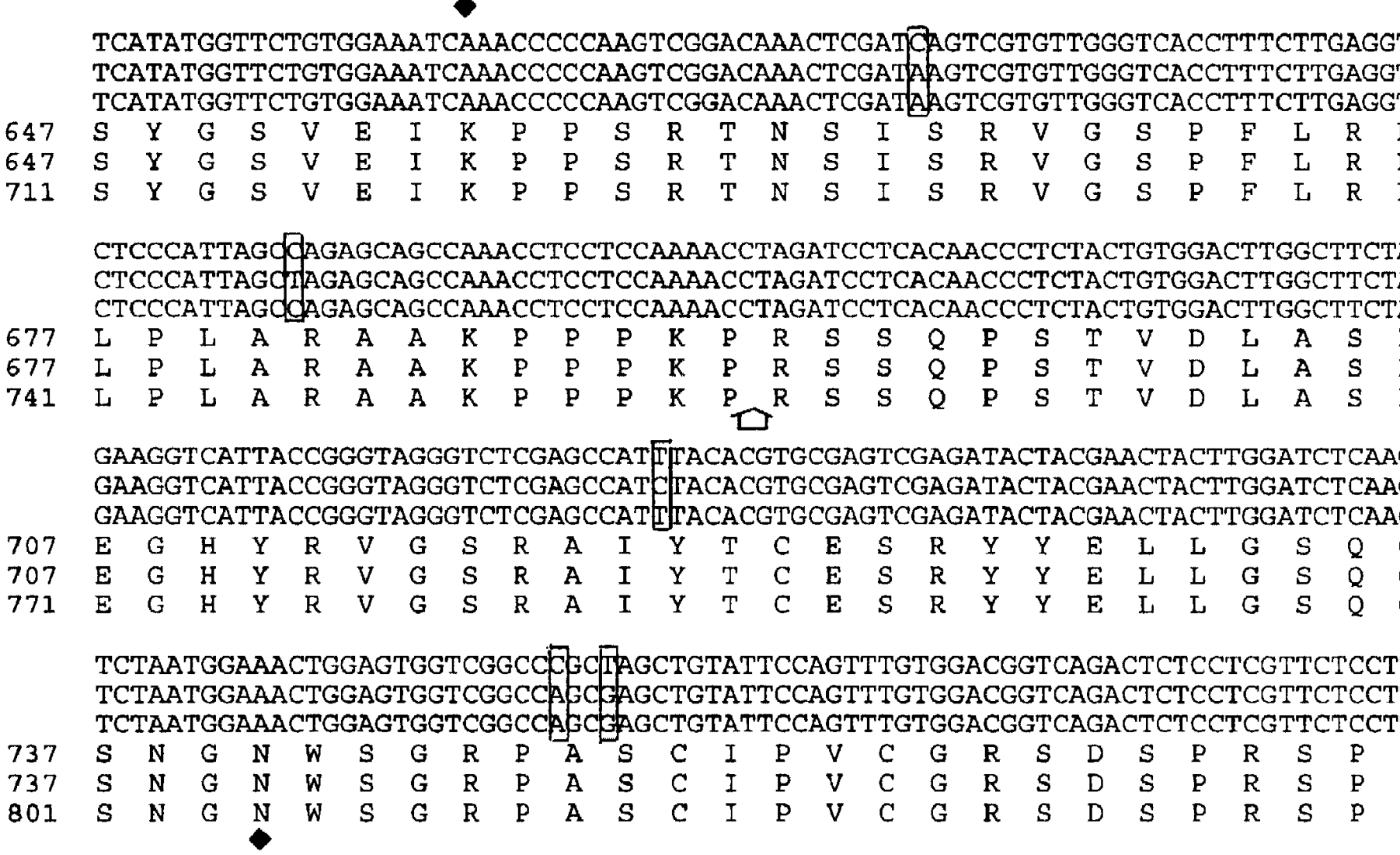


FIG. 11F

Seikagaku Corporation Exhibit 2004 - Page 246 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

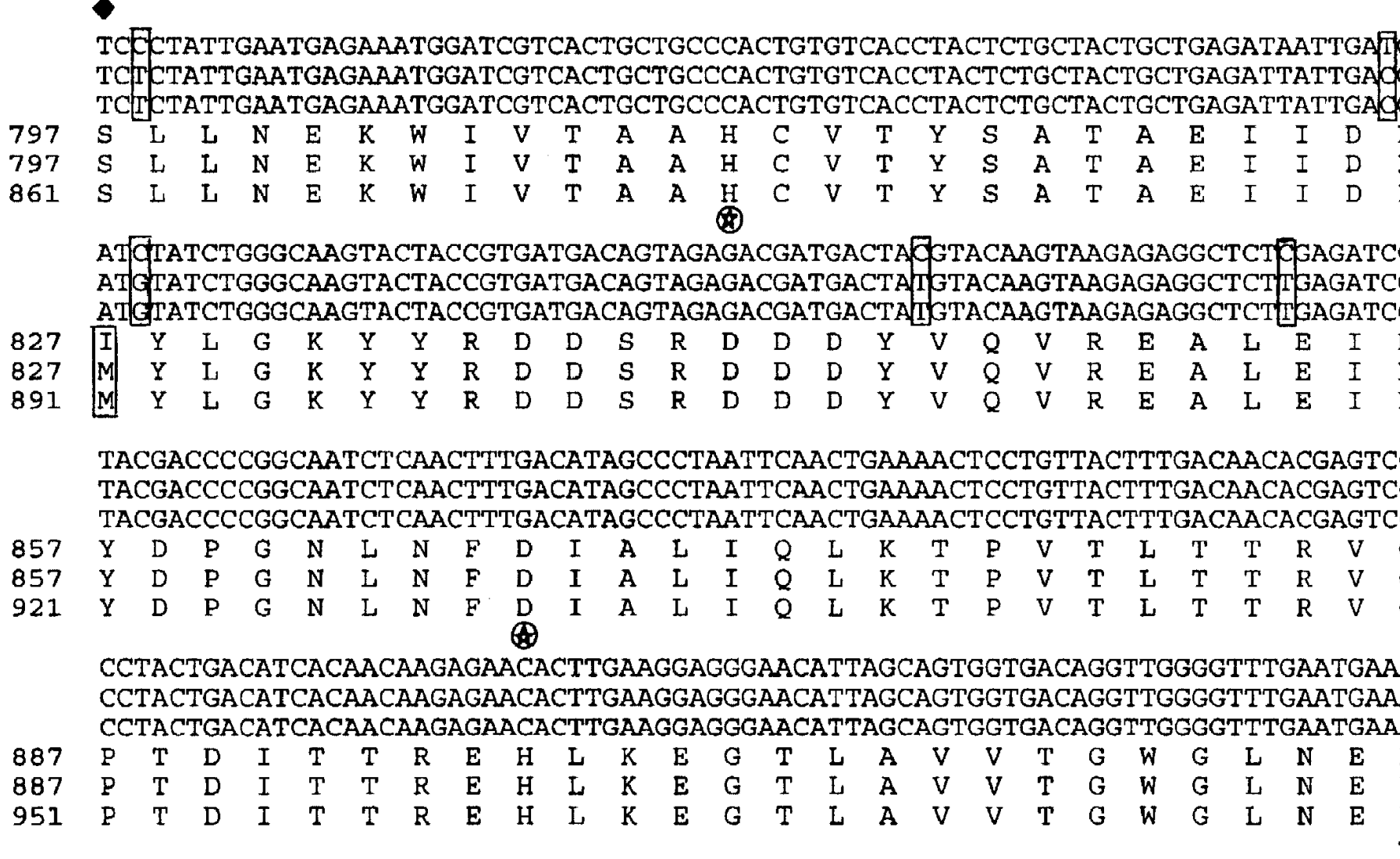


FIG. 11G

Seikagaku Corporation Exhibit 2004 - Page 247 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

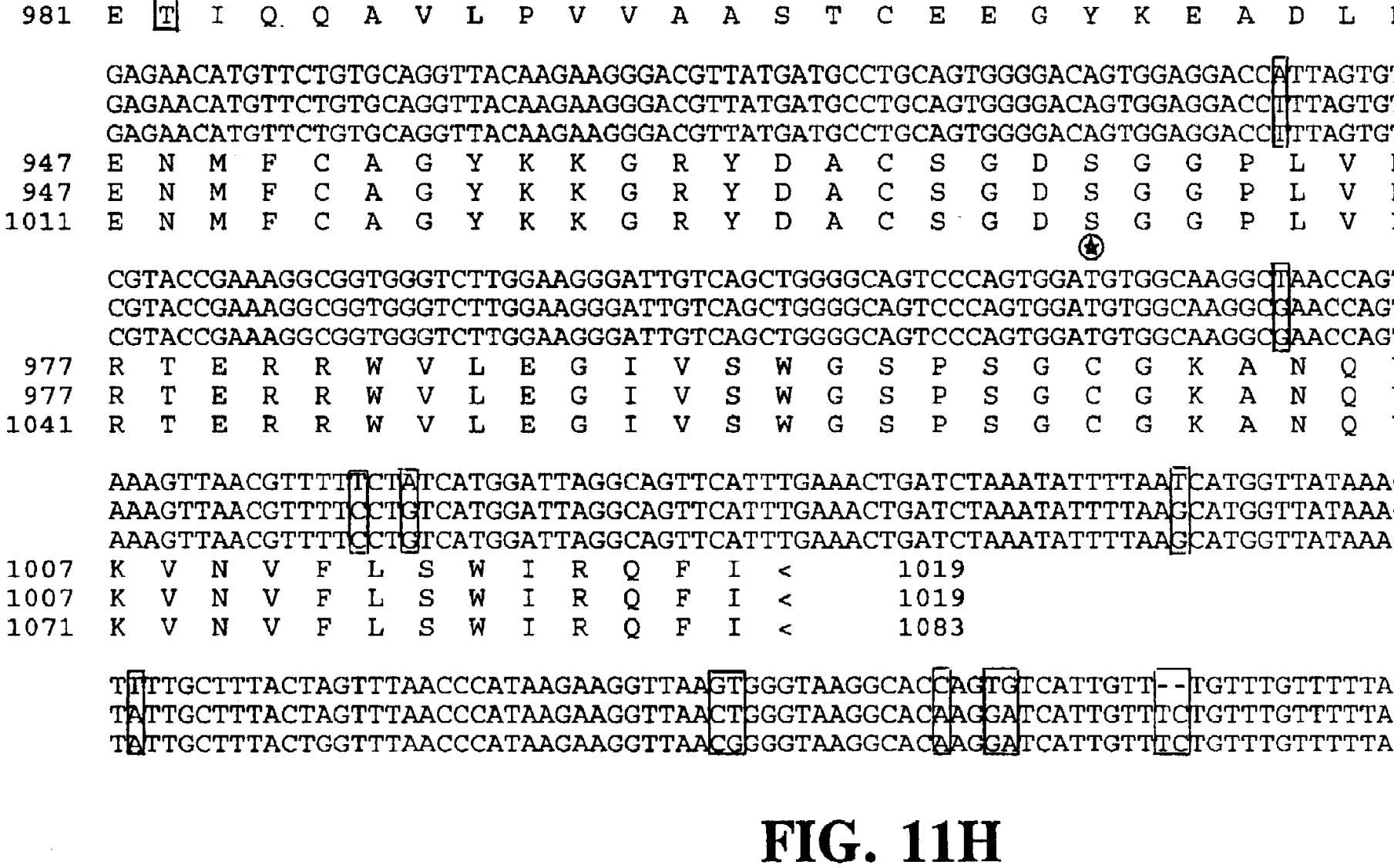
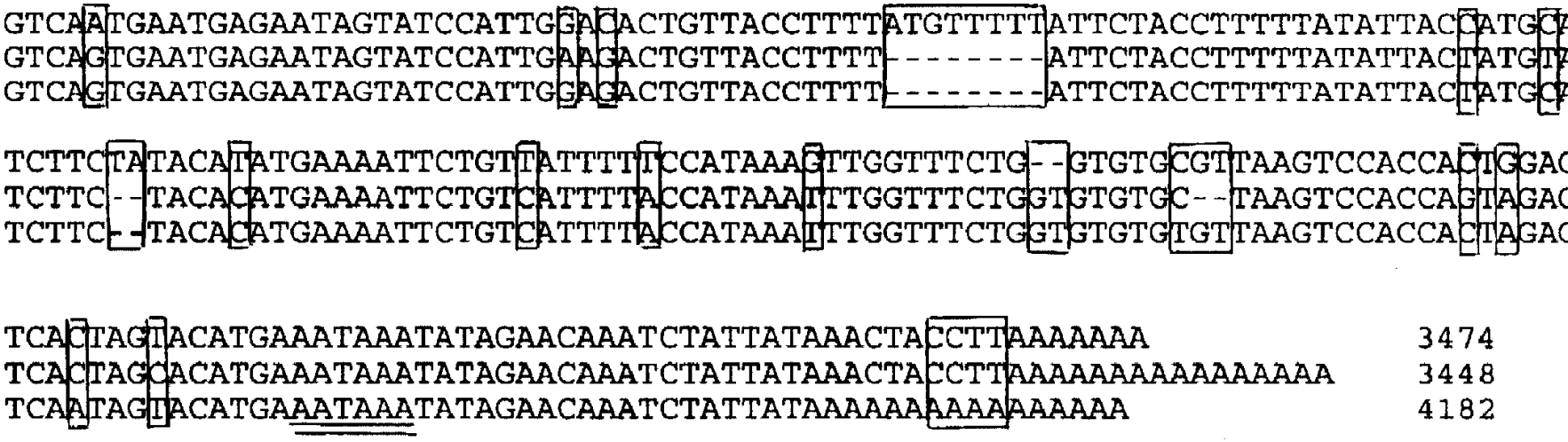


FIG. 11H

Seikagaku Corporation Exhibit 2004 - Page 248 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023



```
GTCAATGAATGAGAATAGTATCCATTGGACACTGTTACCTTTTATGTTTTTATTCTACCTTTTTATATTACCATGCA
GTCAGTGAATGAGAATAGTATCCATTGAAGACTGTTACCTTTT--------ATTCTACCTTTTTATATTACTATGTA
GTCAGTGAATGAGAATAGTATCCATTGGAGACTGTTACCTTTT--------ATTCTACCTTTTTATATTACTATGCA

TCTTCTATACATATGAAAATTCTGTTATTTTTCCATAAAGTTGGTTTCTG--GTGTGCGTTAAGTCCACCACTGGAC
TCTTC--TACACATGAAAATTCTGTCATTTTACCATAAATTTGGTTTCTGGTGTGTGC--TAAGTCCACCAGTAGAC
TCTTC--TACACATGAAAATTCTGTCATTTTACCATAAATTTGGTTTCTGGTGTGTGTGTTAAGTCCACCACTAGAC

TCACTAGTACATGAAATAAATATAGAACAAATCTATTATAAACTACCTTAAAAAAA                3474
TCACTAGCACATGAAATAAATATAGAACAAATCTATTATAAACTACCTTAAAAAAAAAAAAAAAA       3448
TCAATAGTACATGAAATAAATATAGAACAAATCTATTATAAAAAAAAAAAAAAA                  4182
```

FIG. 11I

Seikagaku Corporation Exhibit 2004 - Page 249 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Averages and values plotted are for that base plus the next 6

residue	Score	Avg.	Hydrophilic 0 Hydrophobic
1 M	1.90	0.900	
2 W	-0.90	0.529	
3 V	4.20	1.057	
4 T	-0.70	1.000	
5 C	2.50	1.500	
6 F	2.80	1.743	
7 D	-3.50	1.700	
8 T	-0.70	1.700	
9 F	2.80	1.686	
10 L	3.80	1.171	
11 F	2.80	1.229	
12 V	4.20	1.229	
13 C	2.50	0.986	
14 E	-3.50	1.171	
15 S	-0.80	2.214	
16 S	-0.80	2.686	
17 V	4.20	3.400	
18 F	2.80	2.671	
19 C	2.50	1.629	
20 L	3.80	1.671	
21 L	3.80	1.071	
22 C	2.50	0.929	
23 V	4.20	0.929	
24 W	-0.90	-0.314	
25 R	-4.50	-0.314	
26 F	2.80	-0.314	
27 G	-0.40	-0.114	
28 F	2.80	0.343	
29 C	2.50	-0.243	
30 R	-4.50	-0.714	
31 W	-0.90	0.329	
32 R	-4.50	0.229	
33 V	4.20	1.271	
34 F	2.80	1.271	
35 Y	-1.30	0.314	
36 S	-0.80	0.386	
37 F	2.80	0.400	
38 P	-1.60	0.600	
39 F	2.80	1.429	
40 V	4.20	1.571	
41 K	-3.90	1.514	
42 S	-0.80	1.571	
43 T	-0.70	2.043	
44 V	4.20	1.957	
45 V	4.20	0.900	
46 L	3.80	0.114	
47 L	3.80	-0.543	
48 Q	-3.50	-0.543	
49 C	2.50	-0.500	
50 Y	-1.30	-1.357	
51 H	-3.20	-1.271	
52 Y	-1.30	-0.929	
53 S	-0.80	-1.300	
54 L	3.80	-0.786	
55 H	-3.20	-1.514	
56 N	-3.50	-1.171	
57 T	-0.70	-0.071	
58 S	-0.80	-0.471	
59 K	-3.90	-0.586	

Fig. 12

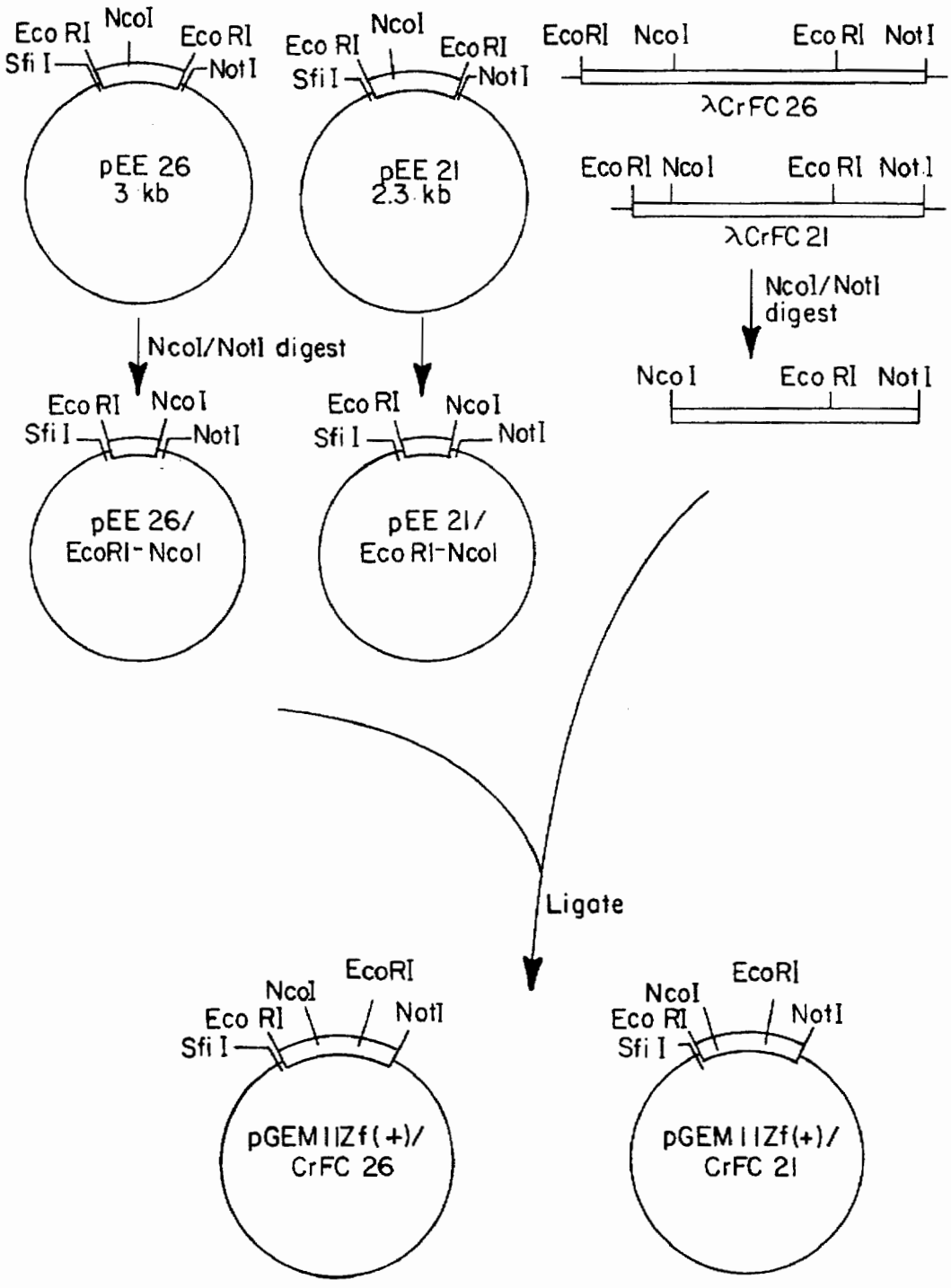

Eco RI
NcoI
Eco RI
Sfi I
NotI
pEE 26
3 kb
pEE 21
2.3 kb
EcoRI NcoI
Eco RI NotI
λCrFC 26
λCrFC 21
NcoI/NotI digest
pEE 26/
EcoRI-NcoI
pEE 21/
Eco RI-NcoI
Ligate
EcoRI
pGEM11Zf(+)/
CrFC 26
pGEM11Zf(+)/
CrFC 21

FIG. 13

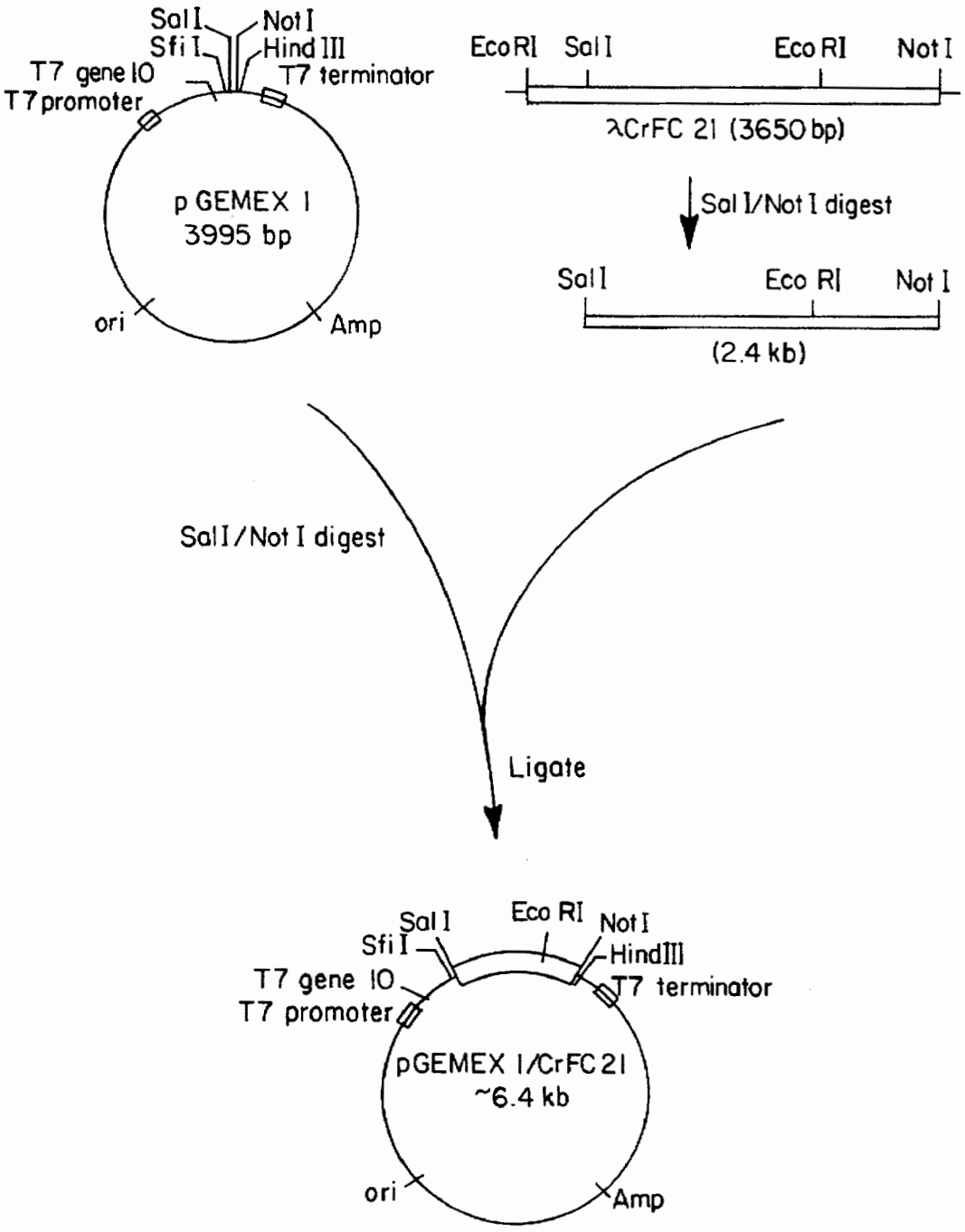

Sal I
Not I
Sfi I
Hind III
T7 gene 10
T7 terminator
T7 promoter
p GEMEX 1
3995 bp
ori
Amp
Eco RI
Sal I
Eco RI
Not I
λCrFC 21 (3650 bp)
Sal I/Not I digest
Sal I
Eco RI
Not I
(2.4 kb)
Sal I/Not I digest
Ligate
Sal I
Eco RI
Not I
Sfi I
HindIII
T7 gene 10
T7 terminator
T7 promoter
pGEMEX 1/CrFC 21
~6.4 kb
ori
Amp

FIG. 14

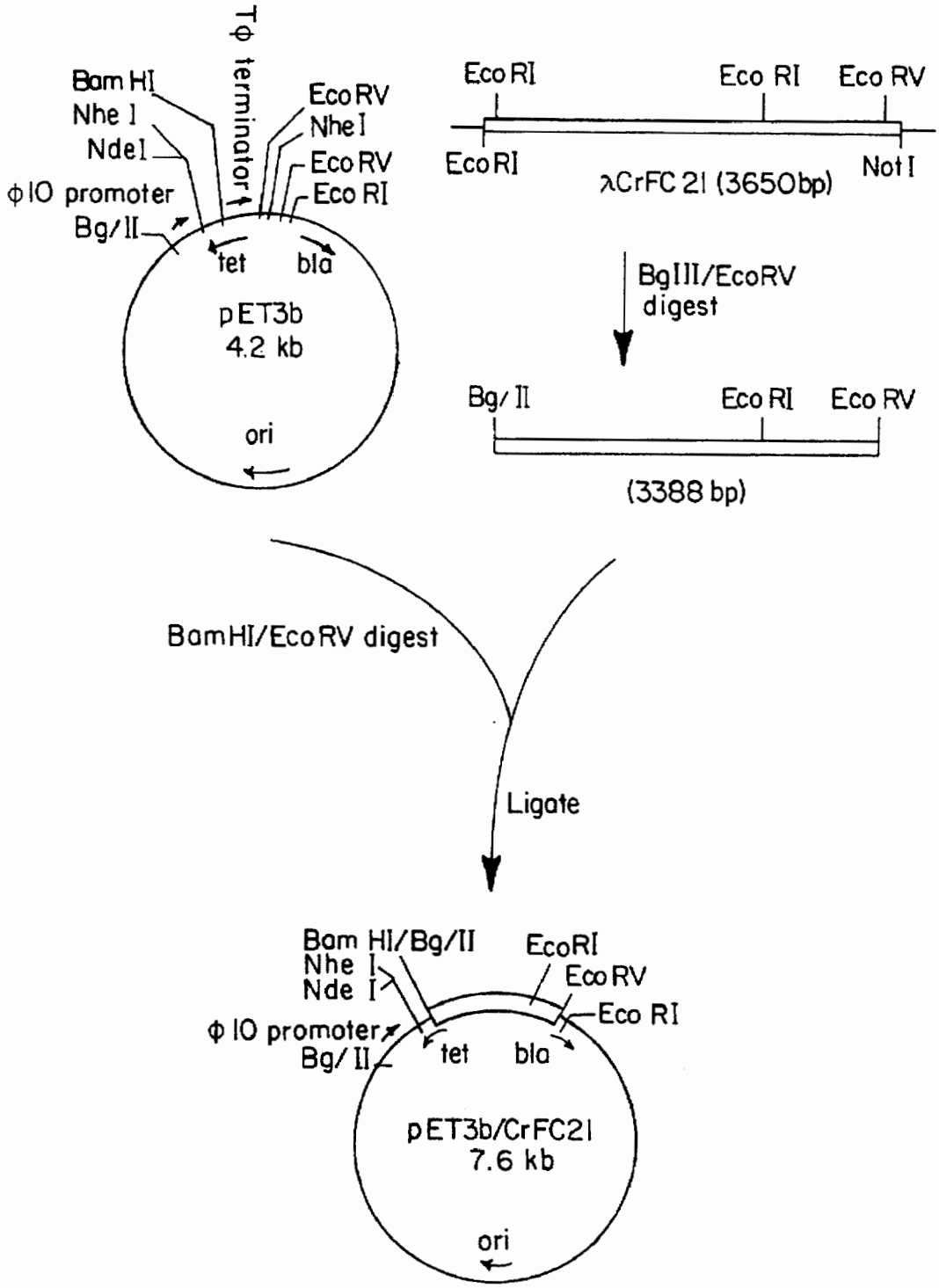

Tφ terminator
Bam HI
Nhe I
Nde I
φ10 promoter
Bg/II
Eco RV
Nhe I
Eco RV
Eco RI
tet
bla
pET3b
4.2 kb
ori
Eco RI
Eco RI
Eco RV
Eco RI
Not I
λCrFC 21 (3650bp)
BgIII/EcoRV
digest
Bg/II
Eco RI
Eco RV
(3388 bp)
BamHI/EcoRV digest
Ligate
Bam HI/Bg/II
Nhe I
Nde I
φ10 promoter
Bg/II
EcoRI
Eco RV
Eco RI
tet
bla
pET3b/CrFC21
7.6 kb
ori

FIG. 15

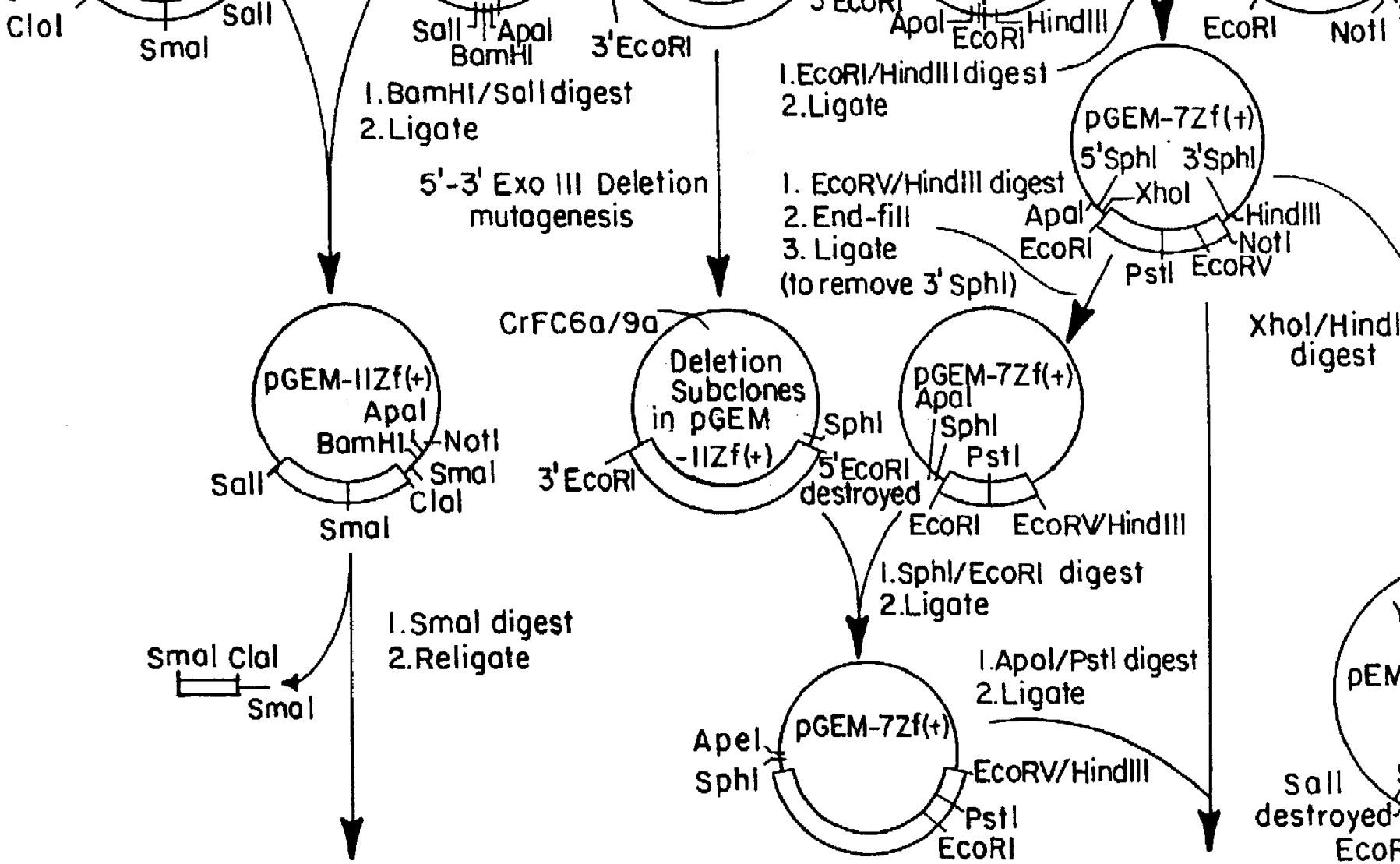


FIG. 16A

Seikagaku Corporation Exhibit 2004 - Page 254 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

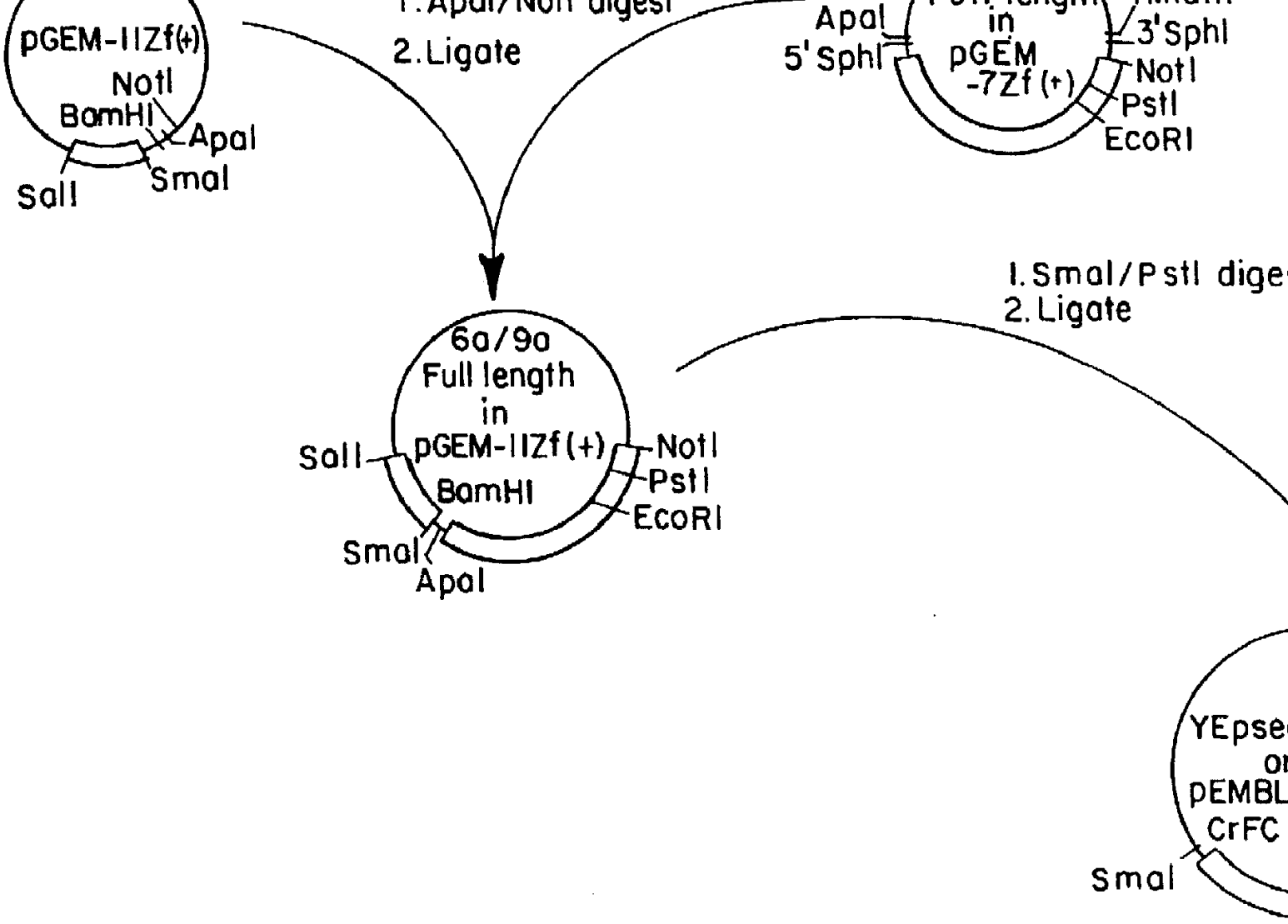


FIG. 16B

Seikagaku Corporation Exhibit 2004 - Page 255 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

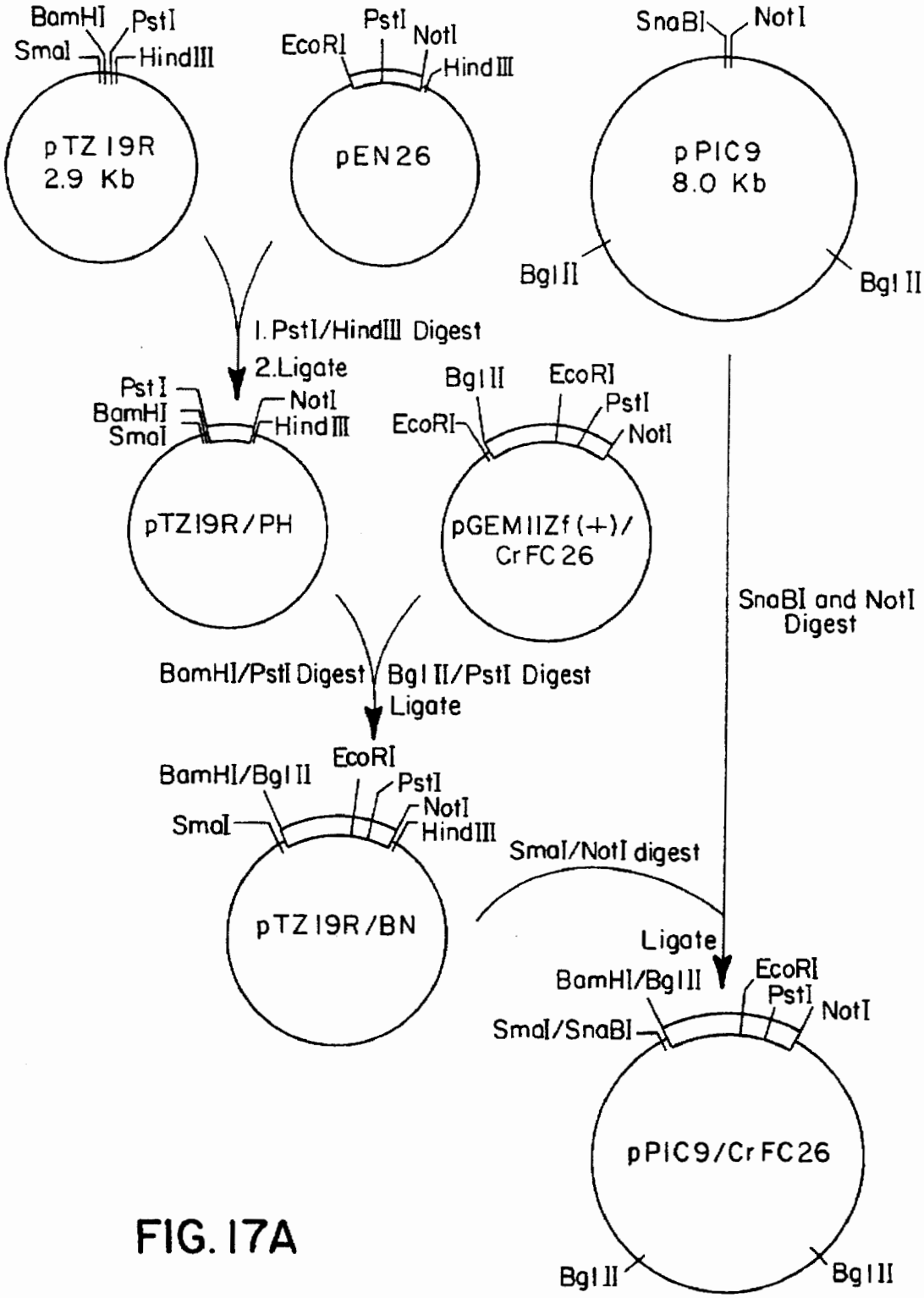

BamHI
PstI
SmaI
HindIII
pTZ 19R
2.9 Kb
EcoRI
PstI
NotI
HindIII
pEN26
SnaBI
NotI
pPIC9
8.0 Kb
BglII
BglII
1. PstI/HindIII Digest
2.Ligate
PstI
BamHI
SmaI
NotI
HindIII
pTZ19R/PH
BglII
EcoRI
PstI
EcoRI
NotI
pGEM11Zf(+)/
CrFC26
SnaBI and NotI
Digest
BamHI/PstI Digest
BglII/PstI Digest
Ligate
BamHI/BglII
EcoRI
PstI
NotI
SmaI
HindIII
pTZ19R/BN
SmaI/NotI digest
Ligate
BamHI/BglII
EcoRI
PstI
NotI
SmaI/SnaBI
pPIC9/CrFC26
BglII
BglII

FIG. 17A

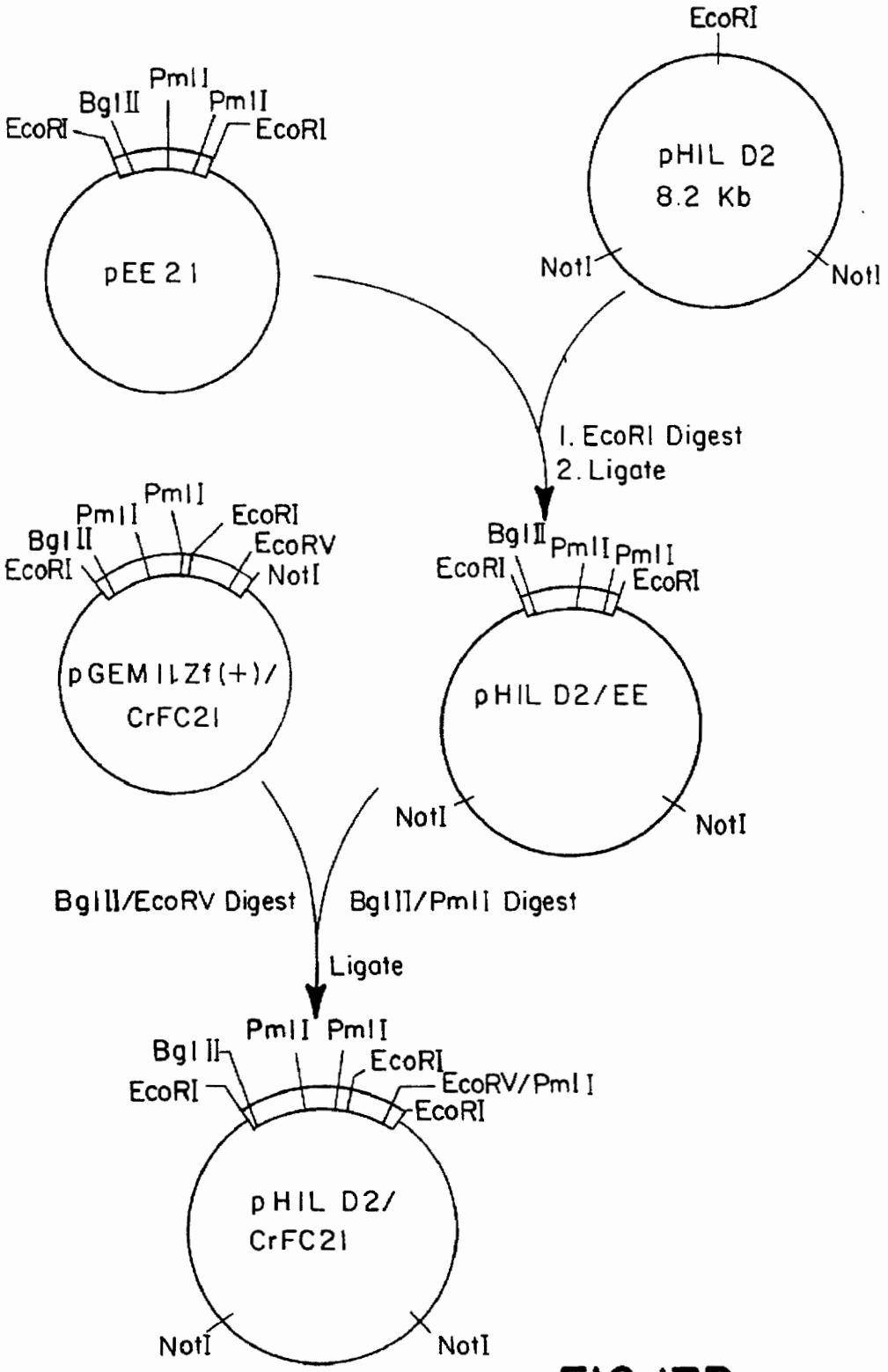

EcoRI
pHIL D2
8.2 Kb
NotI
NotI
PmlI
BglII
PmlI
EcoRI
EcoRI
pEE 21
1. EcoRI Digest
2. Ligate
PmlI
PmlI
EcoRI
BglII
EcoRV
EcoRI
NotI
pGEM 11Zf(+)/
CrFC21
BglII
PmlI
PmlI
EcoRI
EcoRI
pHIL D2/EE
NotI
NotI
BglII/EcoRV Digest
BglII/PmlI Digest
Ligate
PmlI
PmlI
BglII
EcoRI
EcoRI
EcoRV/PmlI
EcoRI
pHIL D2/
CrFC21
NotI
NotI

FIG. 17B

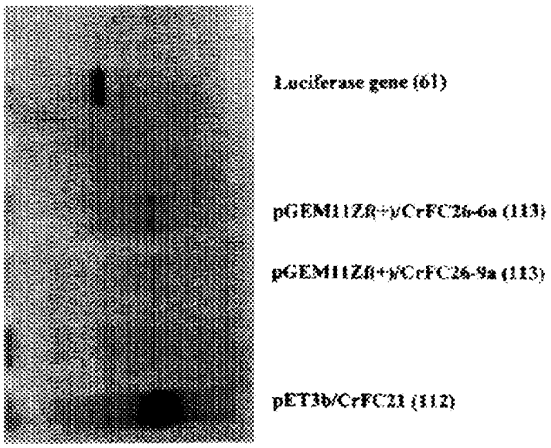


FIG.18A

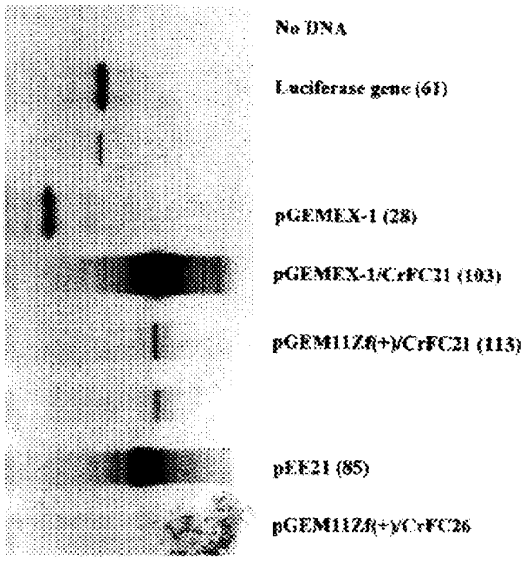


FIG.18B

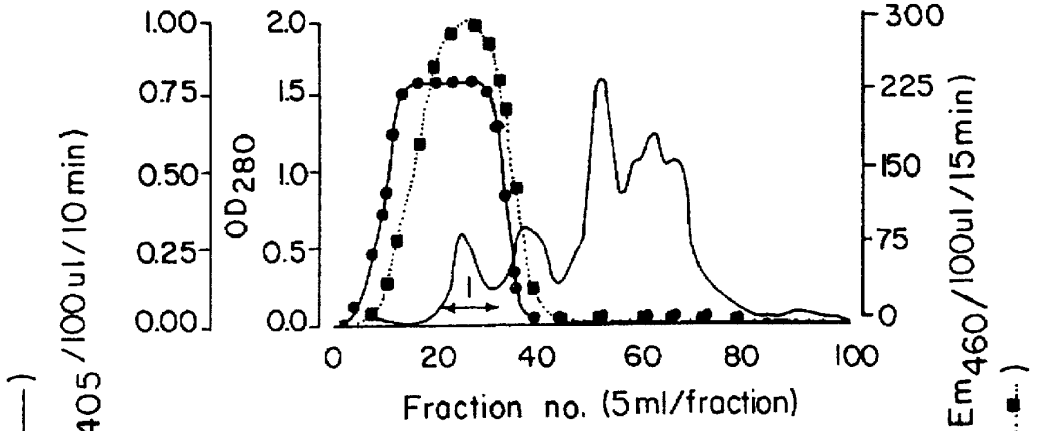


FIG. 19A

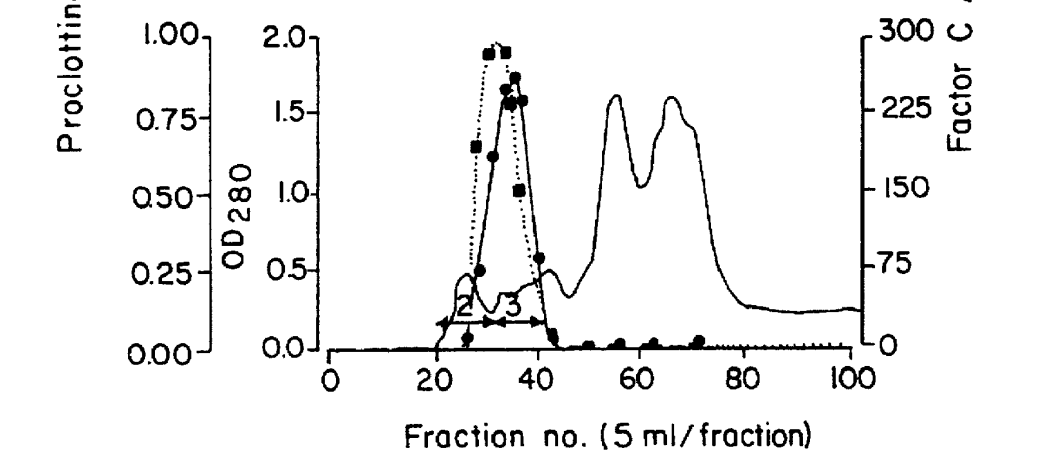


FIG. 19B

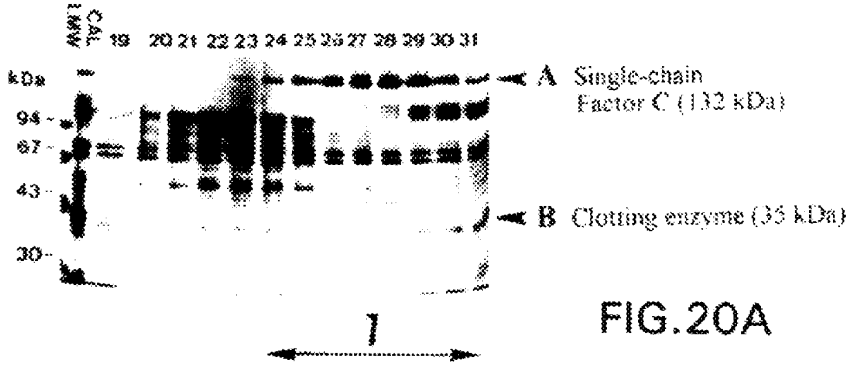


FIG.20A

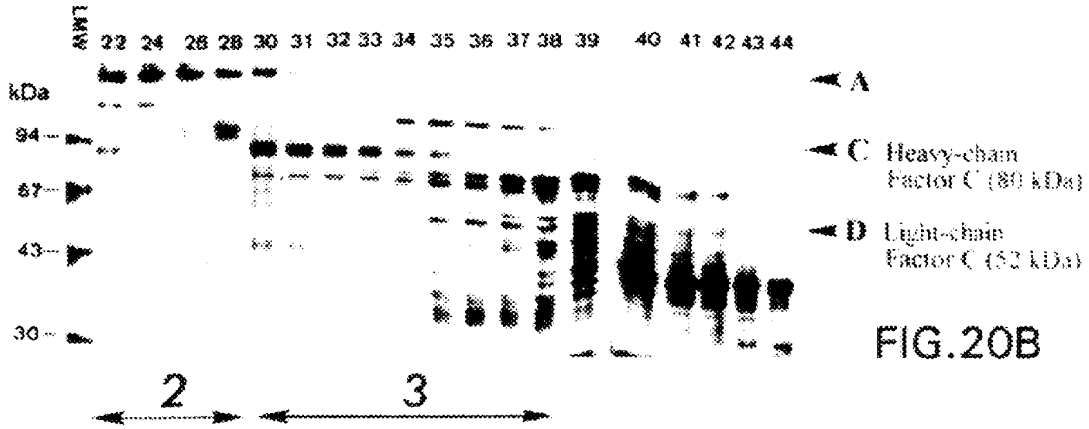


FIG.20B

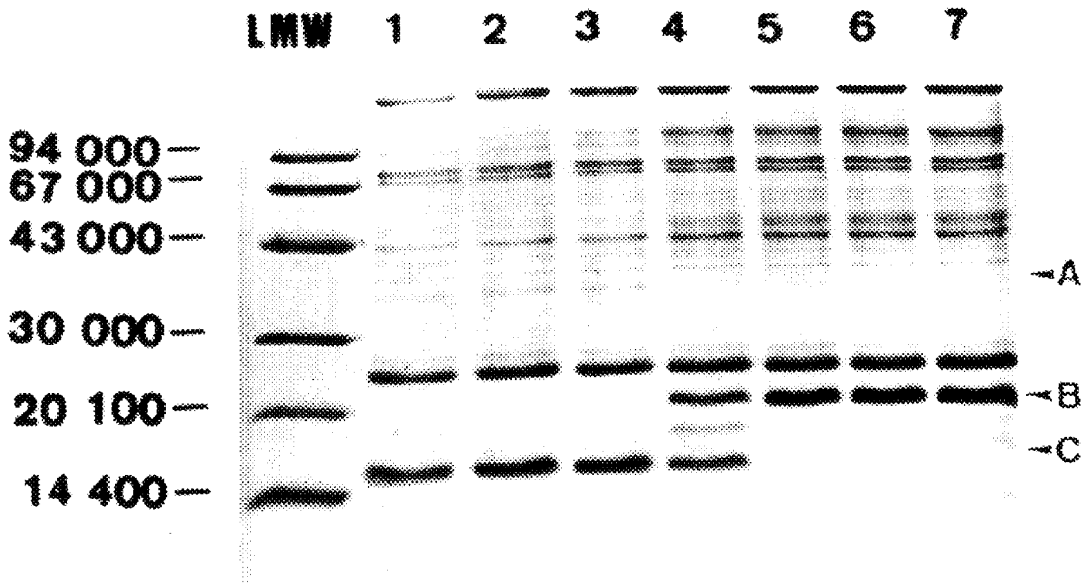


FIG.21

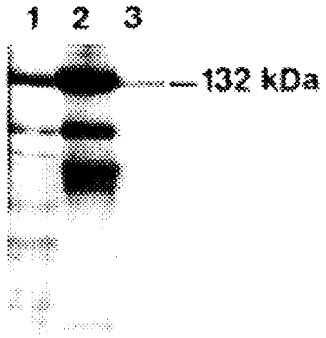


FIG.22A

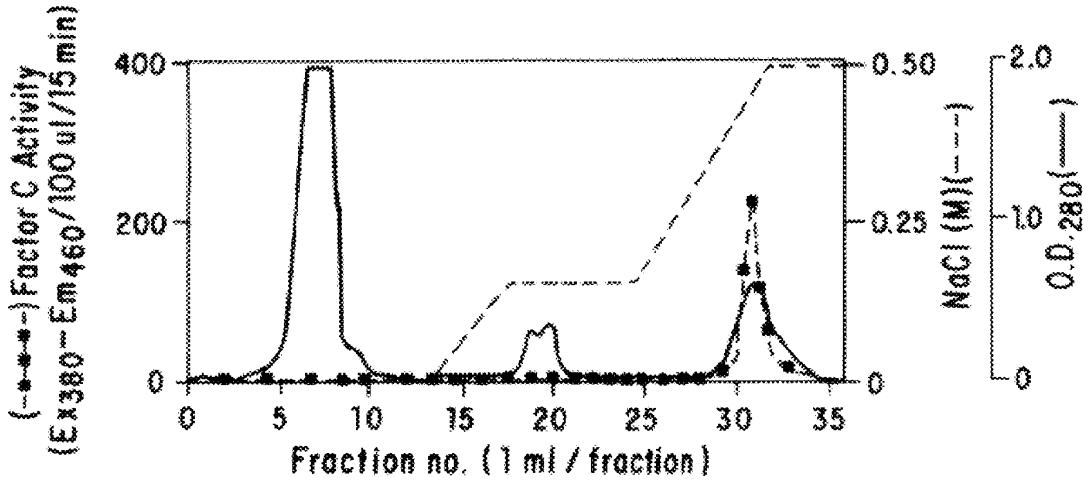


FIG.22B

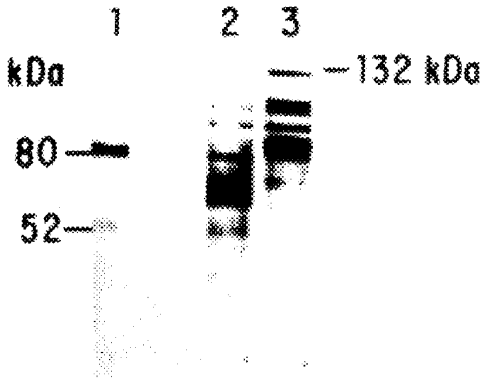


FIG.23A

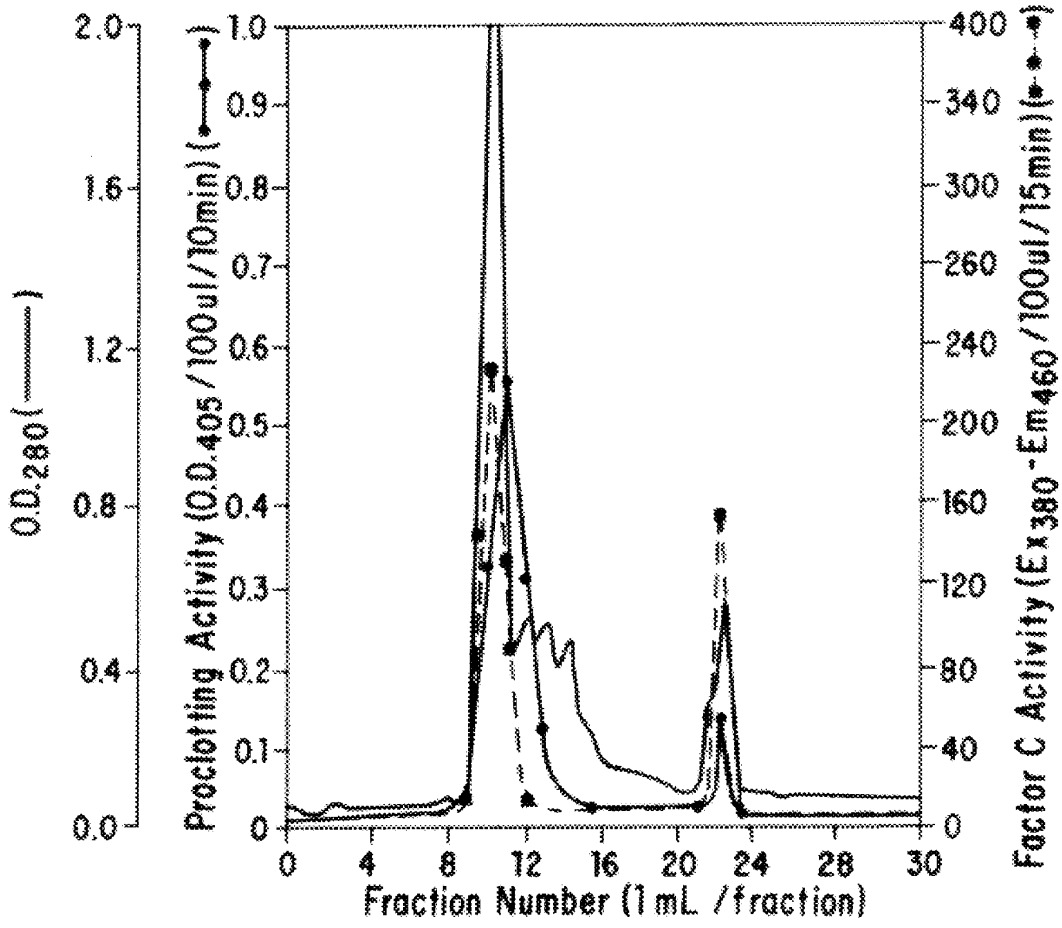


FIG.23B

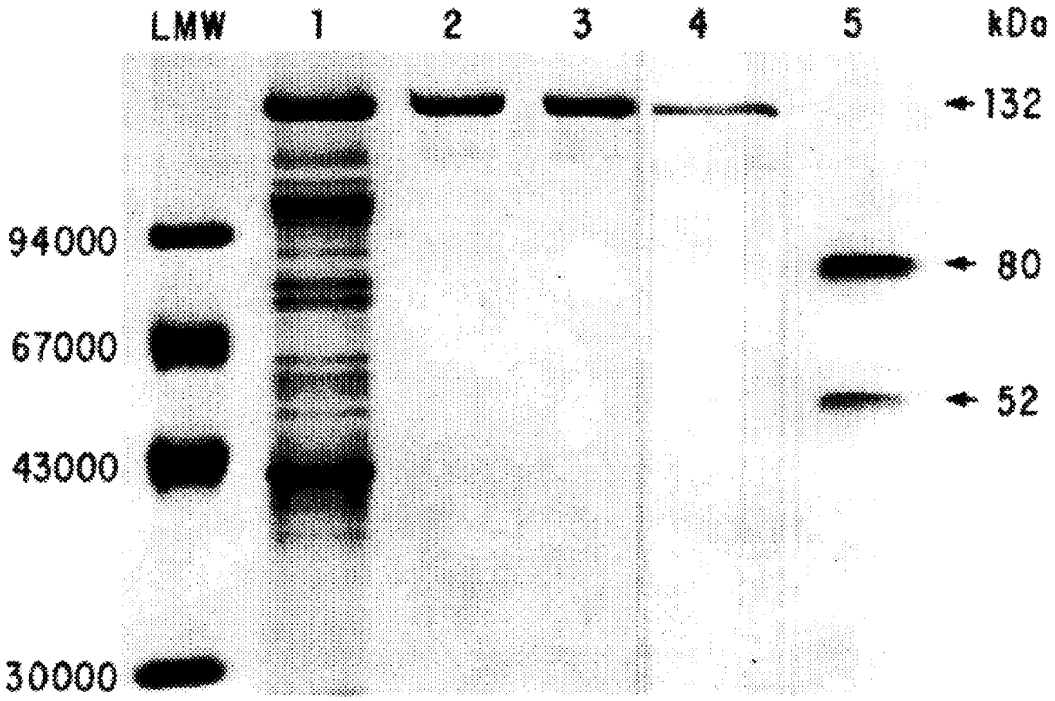


FIG.24

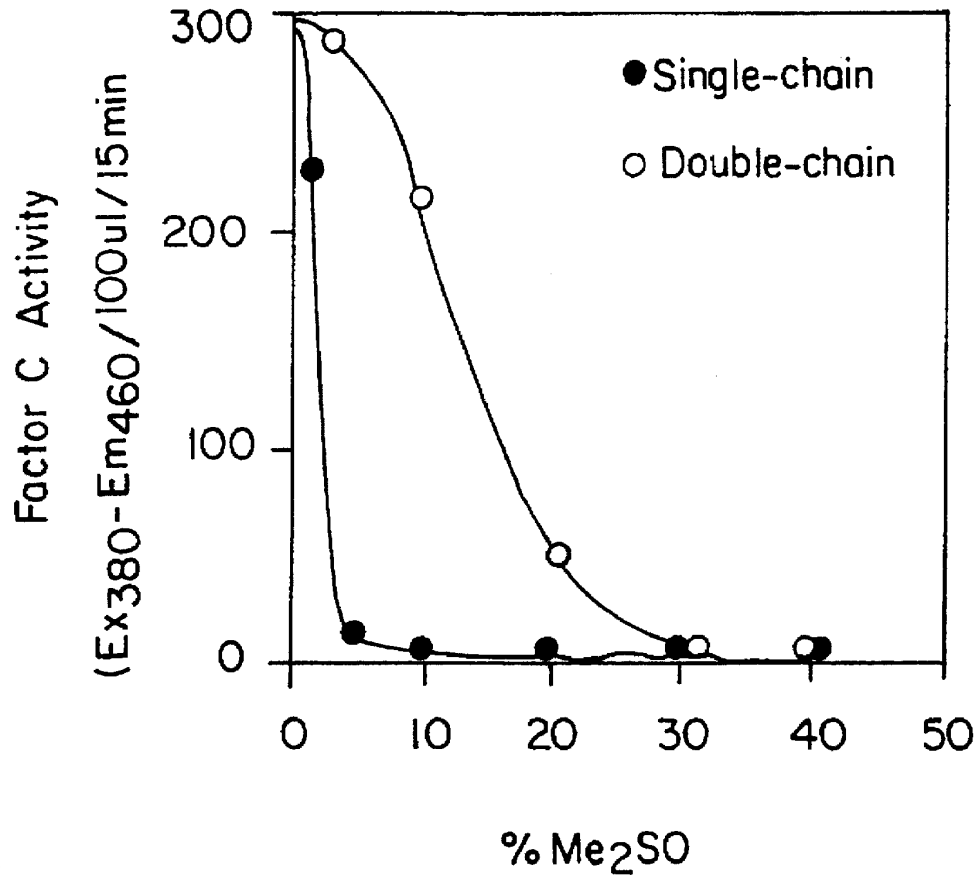

300
200
100
0
Factor C Activity
(Ex380-Em460/100ul/15min
Single-chain
Double-chain
0
10
20
30
40
50
% Me2SO

FIG. 25

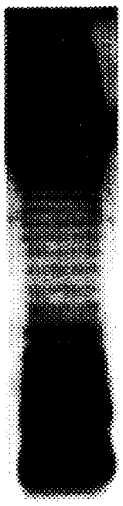


FIG.26A

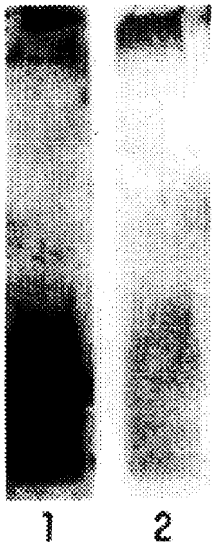


FIG.26B

CLONED FACTOR C CDNA OF THE SINGAPORE HORSESHOE CRAB, *CARCINOSCORPIUS ROTUNDICAUDA* AND PURIFICATION OF FACTOR C PROENZYME

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to cloned cDNAs of *Carcinoscorpius rotundicauda* Factor C useful in the production of Factor C for endotoxin assays and as probes for detecting Factor C genes in other genera and species, Factor C proteins per se, the purification of Factor C proenzyme, and methods for maintaining Factor C, either naturally-occurring or produced via recombinant methods, in its zymogen form.

2. Description of Related Art

There are four extant species of horseshoe crabs: *Limulus polyphemus, Tachypleus tridentatus, Tachypleus gigas* and *Carcinoscorpius rotundicauda*. Within the predominant type of blood cell, i.e., amoebocytes, are enzymes which are activated in the presence of Gram negative bacterial endotoxin, resulting in the final conversion of soluble coagulogen to coagulin clot (FIG. 1). This cascade of enzymes and coagulogen constitutes the commercially available amoebocyte lysate employed for detection of endotoxin. The latter is extremely ubiquitous and indomitable, contaminating parenteral preparations, water, food and pharmaceutical products. The endotoxin also forms the basis of the detection of some Gram negative bacterial infections such as gonorrhoea and meningitis. The Limulus amoebocyte lysate (LAL) constitutes a most rapid and sensitive in vitro assay for detection of femtogram levels of endotoxin (Ho. B, 1983. *Microbios Letts.* 24, 81–84). This diagnostic test for endotoxin forms a crucial FDA-approved pyrogen test which is an integral aspect of quality assurance of many pharmaceutical preparations, especially injectables/parenterals.

Although the production and application of the LAL has become more standardized in recent years (Associates of Cape Cod, Woodshole, Mass., USA, and M. A. Bioproducts, Walkersville, Md., USA), significant variations occur in lysates produced by different manufacturers and even from lot-to-lot within batches produced by individual manufacturers (Ho et al., 1993. *Biochem. & Mol. Biol. Intl.* 29 [4], 687–694). Although the U.S. Limulus population appears unaffected by commercialization, their number could be diminished by overutilization and deterioration of habitat. The Japanese *T. tridentatus* has been pronounced an endangered species and is on its way to extinction (Sekiguchi, K. & Nakamura, K. 1979. In: *Biomedical Applications of the Horseshoe Crabs (Limulidae)*, E. Cohen et al., eds., pp. 37–45, Alan R. Liss, Inc., New York). The availability of a second generation genetically-engineered lysate enzyme receptive to endotoxin, viz., Factor C, would alleviate problems of batch-variation, and also provide a standardized and continuous supply of material for endotoxin/pyrogen assays. This may be achieved through recombinant DNA technology.

Interest in the cloning of Factor C is not new. Japanese workers (Muta et al., 1991, *J. Biol. Chem.* 266, 6554–6561) have cloned the *T. tridentatus* Factor C gene. However, the *T. tridentatus* Factor C gene was cloned in two partial overlapping fragments in separate recombinants (pFC 41 and pFC 53) and reported as a composite DNA sequence totalling 3474 bp.

The potential applications of the genetically-engineered Factor C gene lies in the many possible ways of manipulating and subcloning this gene into a variety of vectors of choice in order to achieve optimum levels of expression of the recombinant Factor C. The biotechnological implications of the resulting recombinant lysate enzyme cannot be overemphasized. Factor C is, afterall, the first enzyme in the amoebocyte lysate coagulation cascade which is activated by endotoxin, which it detects. The recombinant Factor C enzyme may be employed in a chromogenic assay. Upon activation by endotoxin, Factor C converts the substrate to a colored product, thereby detecting and quantifying the endotoxin.

SUMMARY OF THE INVENTION

Accordingly, it is an object of the present invention to provide an isolated, purified DNA molecule comprising a nucleotide sequence that encodes a protein having the same enzymatic activity as Factor C protein in assays for Gram negative bacterial endotoxin. The Factor C protein can be a *Carcinoscorpius rotundicauda* Factor C protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4. Said DNA molecule can comprise a nucleotide sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3, or can comprise a nucleotide sequence selected from the group consisting of a nucleotide sequence that hybridizes to a DNA molecule encoding amino acid sequence SEQ ID NO:2 or SEQ ID NO:4 under salt and temperature conditions equivalent to 5×SSC and 42° C. and that codes on expression for a protein that has the same enzymatic activity as Factor C protein in assays for detecting Gram negative bacterial endotoxin, a nucleotide sequence that is functionally equivalent to a DNA molecule encoding amino acid SEQ ID NO:2 or SEQ ID NO:4 due to the degeneracy of the genetic code and that codes on expression for a protein that has the same enzymatic activity as Factor C protein in assays for detecting Gram negative bacterial endotoxin, and a nucleotide sequence that is functionally equivalent to a DNA molecule encoding amino acid SEQ ID NO:2 or SEQ ID NO:4 in that it codes on expression for a protein in which one or more amino acids has or have been added, deleted, or substituted, but which has the same enzymatic activity as Factor C protein in assays for detecting Gram negative bacterial endotoxin. Said DNA molecule can also comprise the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 1 to 568 have been deleted.

Another object of the present invention is to provide a recombinant vector comprising any of the aforementioned DNA molecules. The vector portion can be a member selected from the group consisting of λgt 22, pGEM 11Zf (+), pGEMEX-1, pET 3b, YEpsec 1, pEMBLyex 4, pIC 9, and pHIL D2.

Another object of the present invention is to provide a host cell transformed with said recombinant vector. Said host cell can be selected from the group consisting of bacteriophage λ, Baculovirus, *E. coli*, a mammalian cell, and a yeast.

Another object of the present invention is to provide a method of producing a protein having the same enzymatic activity as Factor C protein in assays for Gram negative bacterial endotoxin, comprising culturing said host cell under conditions in which said DNA molecule is expressed, and recovering said protein.

Another object of the present invention is to provide an isolated, purified protein molecule comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, a protein molecule comprising an

amino acid sequence that is functionally equivalent to SEQ ID NO:2, in which one or more amino acids has or have been added, deleted, or substituted, but which has the same enzymatic activity as Factor C protein in assays for detecting Gram negative bacterial endotoxin, and a protein molecule comprising an amino acid sequence that is functionally equivalent to SEQ ID NO:4, in which one or more amino acids has or have been added, deleted, or substituted, but which has the same enzymatic activity as Factor C protein in assays for detecting Gram negative bacterial endotoxin.

A further object of the present invention is to provide a process for purifying Factor C zymogens, comprising the steps of:

(a) providing an amoebocyte lysate;

(b) fractionating said amoebocyte lysate of step (a) by affinity chromatography employing a buffer containing dimethylsulfoxide and a chelating agent, and pooling fractions for each peak;

(c) desalting said fractions of step (b) by gel filtration chromatography, and pooling fractions for each peak;

(d) fractionating said fractions of step (c) by affinity chromatography, to obtain a fraction containing purified single-chain Factor C and fractions containing double-chain Factor C;

(e) recovering said purified single-chain Factor C;

(f) further fractionating said fractions of step (d) containing double-chain Factor C by gel filtration chromatography to obtain a fraction containing purified double-chain Factor C; and

(g) recovering said purified double-chain Factor C.

A further object of the present invention is to provide a method for maintaining Factor C in its zymogen form in a crude amoebocyte lysate, comprising lysing amoebocytes in a solution comprising dimethylsulfoxide and, optionally, a chelating agent.

Yet a further object of the present invention is to provide a method for maintaining Factor C expressed by transformed host cells grown in a culture medium in its zymogen form, comprising contacting said Factor C with dimethylsulfoxide and, optionally, a chelating agent, and subsequently isolating said Factor C in the presence of dimethylsulfoxide and, optionally, a chelating agent. Said Factor C can be accumulated intracellularly within said host cells, and said contacting can be performed by lysing said host cells in the presence of dimethyl-sulfoxide and, optionally, a chelating agent. Said Factor C can also be secreted into said culture medium, and said contacting can be performed by adding dimethylsulfoxide and, optionally, a chelating agent to said culture medium prior to isolating said Factor C.

A still further object of the present invention is to provide a method for maintaining Factor C in its zymogen form, comprising contacting said Factor C with dimethylsulfoxide.

A final object of the present invention is the use of the DNA molecules disclosed herein in recombinant processes to produce proteins having the same enzymatic activity as Factor C protein in assays for Gram negative bacterial endotoxin, and as probes for detecting Factor C genes in species other than *Carcinoscorpius rotundicauda*.

Variant forms of Factor C cDNA of the Singaporean estuarine horseshoe crab, *Carcinoscorpius rotundicauda*, have been cloned into the bacteriophage vector λgt 22 and other vectors. These forms have been mapped and sequenced. One of the recombinant clones, λCrFC 26, contains a full-length Factor C cDNA insert of 4182 bp. It includes 568 bp of 5' untranslated sequence containing seven false start ATGs. The open reading frame codes for a signal peptide of 24 amino acids, followed by 1059 residues of the mature Factor C enzyme. There are six potential glycosylation sites and a typical serine protease catalytic triad of Asp-His-Ser in the mature enzyme. The cDNA terminates with 365 bp of 3' untranslated sequence. In comparison with the *Tachypleus tridentatus* Factor C (TtFC) cDNA, there are notable differences in the restriction sites, and subtle base substitutions in the CrFC cDNA. Whereas λCrFC 26 (4182 bp) cDNA has numerous stem-loop structures, thus obscuring its real start codon, another clone, λCrFC 21 (3448 bp) cDNA, has a well-exposed ATG start site. For ease of manipulation, these cDNAs have been recloned into pGEM 11Zf(+). After manipulations of the 5' and 3' ends of the Factor C cDNAs, the major portions of the cDNAs have been subcloned into expression vectors like pGEMEX-1 and pET 3b. The Factor C cDNA has also been recloned into yeast shuttle secretory (YEpsec 1) and non-secretory (pEMBLyex 4) expression vectors. The full-length CrFc 26 and CrFC 21 cDNAs have been excised from their pGEM11Zf(+) vectors and sublconed into Pichia expression vectors pPIC 9 and pHIL-D2.

Using the T_7 promoter in pGEMEX-1 and pET 3b, the CrFC cDNA constructs have been expressed in vitro in the cell-free transcription and translation coupled T_7 expression system.

Further scope of the applicability of the present invention will become apparent from the detailed description and drawings provided below. However, it should be understood that the detailed description and specific examples, while indicating preferred embodiments, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The above and other objects, features, and advantages of the present invention will be better understood from the following detailed descriptions taken in conjunction with the accompanying drawings, all of which are given by way of illustration only, and are not limitative of the present invention, in which:

FIG. 1 shows the coagulation cascade reactions in amoebocytes of *C. rotundicauda*. Endotoxin activates both forms of Factor C. Single-chain Factor C exhibits a reversible activation reaction which signifies a form of feedback regulation in the coagulation cascade (Ding et al., 1993. *Biochim. et Biohphys. Acta*, 1202, 149–156). Double-chain Factor C follows a path previously described in *T. tridentatus* (Iwanaga, S. et al., 1985. In: *Microbiology*, Levie et al., eds., pp. 21–24, Am. Soc. Microbiol., Washington). The broken arrows show an alternative pathway of activation of proclotting enzyme by 1–3 β glucan.

FIG. 2 shows the overall strategy utilized for the synthesis and cloning of the amoebocyte cDNA. Not I primer adaptors were used to prime the first strand synthesis and allow orientation-specific or directional ligation into the vector, bacteriophage λgt 22. A replacement reaction was employed for the second strand synthesis reaction using RNAse H, DNA pol I, and DNA ligase, while Eco RI adaptors were used to produce a sticky ended double-stranded cDNA. After digestion with Not I, the unique restriction ends were phosphorylated with T4 kinase to allow directional ligation to the dephosphorylated Eco RI-Not I digested vector arms. After packaging in vitro, the phage particles were transduced into bacterial host cells (*E. coli* LE392 and Y1090) for subsequent propagation or cloning.

FIG. 3 shows the initial restriction maps of the various Factor C cDNAs isolated and purified from the *C. rotundicauda* amoebocyte cDNA library based on Southern analysis with pFC53 (*T. tridentatus* Factor C cDNA) and the 1100 bp Eco RI-Not I fragment. The cDNAs are arranged in decreasing order of size, and are labelled with the restriction sites Eco RI (E) and Not I (N). Fragments homologous to pFC53 are drawn with solid boxes, while those homologous to the E-N fragment are drawn with engraved boxes.

FIG. 4 shows a comparison between the complete restriction maps of λCrFC 26, λCrFC 21 (*C. rotundicauda*), λFC 53, and λFC 41 (T. tridentatus, see Muta et al., 1991. *J. Biol. Chem.* 266, 6554–6561).

FIG. 5*a* and FIG. 5*b* show the sequencing strategies for clones CrFC 26 and CrFC 21, respectively. Deletion subclones were prepared from pEE and pEN in both directions. The arrows indicate the direction and the extent of the sequences obtained. T_7 promoter primer (⟶) and SP6 promoter primer (⟵) were used.

FIG. 6 shows the complete DNA sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of *C. rotundicauda* Factor C, CrFC 26. The putative signal peptide constitutes the first 24 amino acids. The site of truncation of the signal peptide is indicated by an arrowhead (▲). The potential glycosylation sites are marked with closed diamonds (♦). The amino acid residues constituting the catalytic triad by analogy with trypsin are indicated by (✪). A total of 4 polyadenylation sites (AATAAA) were found in CrFC (two at the 5' noncoding sequence, another within the open reading frame, and the last one at the 3' untranslated region), and each is marked with double underlines (‗). The seven false start ATG sites found upstream of the authentic ATG site are indicated with single underlines (_). These sites are terminated shortly by in-frame stop codons (★) located several bases downstream. The cleavage site of the Factor C enzymes into heavy and light chain intermediates is indicated by a hollow thick arrow (⇦) between residues R and S, while proteolysis of the light chain into A and B chains due to endotoxin activation is indicated by a solid thick arrow (⇦) between the unique phe-ileu site, F and I.

FIG. 7 shows Northern blot analysis of 10 μg total amoebocyte RNA using (a) Eco RI-Eco RI fragment and (b) 369 bp Eco RI-Nde I (5' end) of pCrFC 26 as probes. A single band of approximately 4 kNt was deduced in both blots, indicating that the cDNA isolated is full-length, and that the entire 5' end unequivocally belongs to this species of Factor C.

FIG. 8 shows the DNA sequence (SEQ ID NO:3) and deduced amino acid sequence (SEQ ID NO:4) of CrFC 21. The symbols used to denote the potential glycosylation sites, the catalytic triad, the polyadenylation sites, cleavage site of heavy and light chains (between residues R and S), and the proteolytic site of the light chain into A and B chains (between residues F and I) are similar to those described in FIG. 6.

FIG. 9 shows a comparison of the N terminal nucleotide sequences of different Factor C cDNAs isolated from the horseshoe crabs. The clone pFC53 is from *T. tridentatus* (Muta et al., 1991. *J. Biol. Chem.* 266, 6554–6561), while clones CrFC 21 and CrFC 26 are from *C. rotundicauda* (this work). CrFC 21 and pFC53 have identical 5' end sequences, while CrFC 26 has an extra 716 nucleotides at its 5' end.

FIG. 10*a*–10*c* show the computational predictions of the secondary structures of mRNAs of (a) CrFC 21 and (b) CrFC 26. The leader sequence of CrFC 26 shows numerous hairpin stems and loops. The authentic AUG start codon is boxed. The start codon in CrFC 21 is exposed in a loop, and appears more accessible.

FIG. 11 shows the alignment of the complete DNA sequence and deduced amino acid sequence of *C. rotundicauda* Factor C (CrFC 26 and CrFC 21) with the Factor C of *T. tridentatus* (TtFC, adapted from Muta et al. 1991. *J. Biol. Chem.* 266, 6554–6561). The numbering of the amino acid residues is found at the left, while that of the DNA (in bp) is found at the right. The site of truncation of the signal peptide is indicated by a small arrow head (▲). The "start of homology" region is indicated by a directional arrow (↱). The differences in both the DNA and amino acid sequences are boxed (□). The cleavage site of the Factor C enzymes into heavy and light chain intermediates occurs between residues R and S, indicated by a hollow thick arrow (⇦). Proteolysis of the light chain into A and B subunits due to endotoxin activation is indicated by a solid thick arrow (⇦) between F and I. The remaining symbols are as per legend to FIG. 6.

FIG. 12 shows hydropathy analysis of the first 59 deduced amino acid sequence of CrFC 26. The peak spanning residues 1–24 represents the putative signal sequence.

FIG. 13 shows manipulations of pEE 21 (Eco RI-Eco RI=2.3 kb) and pEE 26 (Eco RI-Eco RI=3 kb), and λCrFC 21 and λCrFC 26, to reclone full-length Factor C cDNAs into pGEM11Zf(+). The pEE of both subclones 21 and 26 being originally subcloned in pGEM11Zf(+) (see Example 8) were digested with Nco I and Not I enzymes. In parallel, the original λCrFC 21 and λCrFC 26 recombinants were also digested with Nco I and Not I to release the partial fragments of Factor C cDNA inserts flanked by Nco I-Not I. The pGEM 11Zf(+) containing the remaining Factor C cDNA inserts flanked by Eco RI-Nco I were then ligated to their corresponding Nco I-Not I inserts of Factor C derived from λCrFC clones. Thus, pGEM 11Zf(+)/CrFC 21 and pGEM 11Zf(+)/CrFC 26 were obtained, each containing the complete CrFC cDNAs of clones 21 and 26, respectively. These full length Factor C cDNA inserts could thus be excised intact by digestion with Sfi I and Not I enzymes.

FIG. 14 shows the subcloning of CrFC 21 into pGEMEX-1. The vector pGEMEX-1 codes for a peptide of 260 amino acids which will be fused to the expressed Factor C protein. λCrFC 21 and pGEMEX-1 were digested with Sal I and Not I. The truncated portion of the Factor C cDNA insert of 2.4 kb (flanked by Sal I-Not I) was ligated in-frame into pGEMEX-1. The recombinant pGEMEX-1/CrFC 21 was then transformed into *E. coli* JM 109 (DE 3 lysogenic strain).

FIG. 15 shows the subcloning of CrFC 21 in pET 3b (pAR 3039). The vector, pET 3b, was constructed in such a way that the foreign gene is cloned in three different reading frames relative to the gene10 initiation codon in the Bam HI site (GGA, GAT, or ATC). For CrFC 21, the GAT codon would give the correct frame for expression. To linearize the vector, pET3b was digested with Bam HI and Eco RV. λCrFC 21 was digested with Bgl II and Eco RV. Bam HI and Bgl II have compatible ends. Ligation of the CrFC of 3388 bp (flanked by Bgl II-Eco RV) with the linearised pET 3b resulted in recombinants containing CrFC, the expression of which will be driven by the φ10 promoter to produce a fusion protein linked to the first 11 amino acids of the gene10 protein. The recombinant was transformed into *E. coli* JM109 (DE 3 lysogenic strain). The FC insert can be released by digestion with Bgl II (upstream) and Eco RV (downstream).

Seikagaku Corporation Exhibit 2004 - Page 269 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

FIG. **16** shows the strategy for cloning CrFC 26 cDNA into *S. cerevisiae* vectors, YEpsecl and pEMBLyex4. The EcoRI-EcoRI fragment (pEE26) and EcoRI-NotI fragment (pEN26) of CrFC 26 were first cloned individually into pGEM11Zf(+). The 5' untranslated sequences and varying lengths of the DNA encoding the leader peptide of CrFC 26 were deleted by performing 5'-3' Exo III deletion mutagenesis on the EcoRI-EcoRI fragment. The complete deletion mutants were reconstructed in pGEM7Zf(+) by ligating the 5' deleted EcoRI-EcoRI fragment of CrFC26 to the EcoRI-NotI fragment of CrFC 26. To facilitate subsequent manipulations, the complete deletion mutants (6a and 9a) were cloned into a modified pGEM11Zf(+). This plasmid was constructed by inserting a DNA fragment containing a Sma I site and a stuffer DNA segment (shaded blocks) from pBluescript II SK-into the multiple cloning site of pGEM11Zf(+). The deletion mutants were then excised with Sma I and Pst I and inserted into the secretory (YEpsecl) and non-secretory (pEMBLyex4) yeast expression vectors. Unless otherwise indicated, all DNA inserts are oriented in a 5' (left)-3' (right) direction.

FIG. **17***a* shows the strategy for cloning CrFC 26 into the *P. pastoris* vector pPIC 9. The vector along with the gene of interest (cloned in frame for the secretory signal) was digested with Bgl II to release the AOX1 flanking insert for transformation into the yeast. Therefore, it was necessary to obliterate the Bgl II site in the CrFC gene and to create a 5' blunt end in frame for the Sna B1 site of the vector. The insert was first subcloned into pTZ19R in two steps. The PstI-HindIII fragment of CrFC insert (derived from pEN26) was first introduced into the vector, followed by the insertion of BglII-PstI fragment of CrFC 26 insert (derived from pGEM11Zf(+)/CrFC 26) into the BamHI-PstI cloning sites of pTZ19R/PH. This step removed the internal Bgl II site in CrFC 26. The insert was then excised using Sma I and Not I digestion from the pTZ19R/BN construct for cloning into compatible Sna B1 and Not I sites in pPIC 9. This construct is henceforth referred to as pPIC9/CrFC 26. FIG. **17***b* shows the strategy for cloning CrFC 21 into the *P. pastoris* vector pHIL D2, which is a non-secretory vector. The gene is translated using its own ATG. The EcoRI-EcoRI fragment of Factor C insert, pEE21, was first cloned into the Eco RI site of the pHIL D2 vector. This construct was digested with Bgl II and Pml I to accomodate the BglII-EcoRV fragment of CrFC from the pGEM11Zf(+)/CrFC 21 to generate the full length insert in pHIL D2. This strategy was used to take advantage of the absence of Bgl II site in the pHIL D2 vector. Not I digestion of the vector allowed directional integration of the CrFC gene into the *P. pastoris* vector.

FIG. **18** shows the products of in vitro cell-free expression of CrFC cDNA constructs in plasmid vectors containing T_7 promoters. The Promega in vitro transcription and translation (TnT T7 coupled reticulocyte lysate) system was used to test the expression potentials of 2 μg each of these constructs. ^{35}S-cys was used to label the translation products. The Factor C proteins were either produced as fusion or non-fusion proteins depending on their vectors: pGEMEX-1/CrFC 21 (Factor C insert coding sequence of 2036 bp+780 bp of fusion gene in vector) yielded a Factor C-fusion protein of 103 kDa; pGEM11Zf(+)/CrFC21 (Factor C insert coding sequence of 3083 bp) yielded a protein of 113 kDa; pEE21 (pGEM11Zf(+)/CrFC 21, containing 2300 bp Eco RI-Eco RI fragment of Factor C coding insert) yielded an 85 kDa protein; pGEM11Zf(+)/CrFC 26-6a and -9a (deletion mutants of CrFC 26, harbouring 3421 bp and 3462 bp, respectively, of Factor C insert coding sequence) each yielded limited amounts of Factor C of 113 kDa; pET3b/CrFC 21 (Factor C insert coding sequence of 3023 bp+30 bp fusion gene 10 of the vector) yielded a Factor C-fusion protein of 112 kDa; pGEM11Zf(+)/CrFC 26 (with full-length Factor C insert of 4182 bp, intact with the 5' untranslated region containing the 7 false start codons) did not express any Factor C protein. Positive control: luciferase gene gave the expected protein of 61 kDa. Negative control: pGEMEX-1 vector expressed a 28 kDa fusion protein.

FIG. **19**(*a*) shows the elution profile of whole CAL lysate chromatographed on Sepharose CL-6B (2.6×97 cm) equilibrated with 0.05M Tris-HCl pH 8.0, containing 0.154M NaCl. FIG. **19**(*b*) shows the elution profile with the addition of 5% Me_2SO and 1 mM Na2EDTA. A total of 326 mg CAL was loaded onto the column, which was run at a flow rate of 20 ml/h. Panel (a) shows Factor C (assayed according to Nakamura et al., 1986. *Eur. J. Biochem.* 154, 511–521) and proclotting enzyme (assayed according to Harada-Suzuki et al., 1982. *J. Biochem.* 92, 793–800) co-eluting over a broad range into the first peak. However, panel (b) shows narrower peaks of these enzyme activities. Double-headed arrows (←→) indicate all pooled fractions with corresponding electrophoretic profiles shown in FIG. **20**.

FIG. **20** shows reducing SDS-PAGE (8% polyacrylamide) analysis of protein fractions obtained from the Sepharose CL-6B column using buffers without (a) and with (b) 5% Me_2SO and 1 mM Na_2EDTA. Fraction numbers are written on top of each lane, and double-headed arrows correspond to pooled fractions found in FIG. **19**. A total amount of 15 μg protein was loaded into each lane. Low molecular weight (LMW) marker proteins were used to estimate the sizes of the protein bands of interest. In (a), single-chain Factor C (band A, 132 kDa) was shown to co-elute with the heavy chain of clotting enzyme (band B, 35 kDa). In (b), single-chain Factor C eluted separately (arrow 2) from the other enzymes, although no apparent activity was detected due to the presence of 5% Me_2SO, which inactivates single-chain Factor C activity (also see FIG. **25**). The heavy chain (band C) and light chain (band D) of the double-chain Factor C were found in fractions 30 (▸) onwards as demarcated by double-headed arrow No. 3. These fractions gave high enzyme activity as it was not inhibited by 5% Me_2SO (refer to FIG. **19**). The 52 kDa light-chain (band D) of the double-chain Factor C is glycosylated and is thus less intensely stained by Coomassie blue.

FIG. **21** shows routine preparations of Carcinoscorpius amoebocyte lysate analyzed on SDS-reducing PAGE (15%). The seven batches of lysate preparations were loaded in the PAGE gel in order of decreasing extent of autoactivation. Lanes 1–3 show the presence of the heavy chain of clotting enzyme (band A, 35 kDa), and complete conversion of coagulogen (band B, 21 kDa) to coagulin (band C, 17 kDa). Thus, appearance of bands A and C is indicative of autoactivation and poor lysate quality. On the other hand, lysates in lanes 5–7 are considered to be of high quality since no coagulin is apparent. Lane 4 shows a batch of lysate which is of intermediate quality and sensitivity in its detection of endotoxin.

FIG. **22** shows affinity chromatography of pooled first peak fractions (refer to arrow 2, FIG. **19***b*) on a heparin-Sepharose CL-6B column (1×11 cm). The silver-stained electrophoretic gel profile shows: Lane 1, whole CAL; 2, pooled first peak fractions (arrow 2 in FIG. **19***b* and **20***b*); 3, single-chain Factor C.

FIG. **23** shows the purification of double-chain Factor C by gel filtration on FPLC Superose 12 HR 10/30 column which was equilibrated at room temperature with a flow rate

Seikagaku Corporation Exhibit 2004 - Page 270 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

of 0.4 ml/min. The pooled enzyme fractions (lane 2) equivalent to 1.2 mg protein was loaded into the column. This was previously eluted from the heparin-Sepharose CL-6B column. The double-chain Factor C was eluted in a single tube (fraction 22), and the silver-stained reducing SDS-PAGE gel shows that it consists of 80 kDa and 52 kDa subunits (lane 1). The pooled first peak fractions (refer to arrow 2, FIG. **19***b*) containing the single-chain form is shown in lane 3. Each lane consists of 15 μg total protein.

FIG. **24** shows a comparison of single-chain (lanes 2 and 4) and double-chain (lanes 3 and 5) forms of Factor C electrophoresed in SDS-PAGE in the absence (lanes 2 and 3) and presence (lanes 4 and 5) of β-mercaptoethanol. Lane 1 contained 15 μg of CAL.

FIG. **25** shows the effect of increasing concentrations of Me_2SO on the two forms of purified Factor C. Prior to the addition of the substrate, 5 μg of each enzyme was incubated for 1 h with 5 μg endotoxin. Single-chain Factor C was more susceptible, losing 95% of its activity at 5% Me_2SO, while double-chain Factor C was totally inactivated only at 30% Me_2SO.

FIG. **26** shows endotoxin binding assay of Factor C. Panel A shows silver-stained endotoxin electrophoresed under reducing conditions on a 15% polyacrylamide gel. Panel B shows a modified Western blot where 10 μg of electrophoretically resolved endotoxin was electroblotted onto a PVDF membrane (LPS strips), reacted with single-chain (lane 1) and double-chain Factor C (lane 2) followed by whole CAL antibody and stained with peroxidase-conjugated goat anti-rabbit antibody. The stained portions indicate the lipid A moiety which is known to be the functional component of endotoxins that specifically binds Factor C (Lei and Morrison, 1988. *J. Immunol.* 141, 996–1011). Control LPS-strips, with bound single- or double-chain Factor C incubated with normal rabbit serum did not show any hybridization.

DETAILED DESCRIPTION OF THE INVENTION

The following detailed description of the invention is provided to aid those skilled in the art in practicing the present invention. Even so, the following detailed description should not be construed to unduly limit the present invention, as modifications and variations in the embodiments herein discussed may be made by those of ordinary skill in the art without departing from the spirit or scope of the present inventive discovery.

The contents of each of the references cited herein are herein incorporated by reference in their entirety.

Cloning of the *Carcinoscorpius rotundicauda* Factor C cDNA

EXAMPLE 1

Construction of Factor C cDNA λ Recombinant Clones

Poly $(A)^+$ RNA Preparation

From 2.5 g wet weight of amoebocytes, total RNA was purified by using guanidinium isothiocynate (Chirgwin et al., 1987. *Biochemistry* 18, 6294–5299) and ultracentrifugation through a CsCl gradient. The total cellular RNA was digested with RNase-free DNase I (BRL) and extracted with phenol/chloroform. Poly $(A)^+$ RNA was purified by chromatography through Oligo-dT Tris Acryl (IBF, France).

cDNA Synthesis and Cloning

The mRNAs purified from amoebocytes were used to synthesize cDNAs following a modification of the method of Gubler and Hoffman, 1983 (*Gene* 25, 263–269). Not I primer adaptors were used to prime the first strand cDNA synthesis to ensure orientation-specific ligation to λgt 22 (Promega, USA). A replacement reaction was employed for the second strand synthesis, while Eco RI adaptors were used to produce sticky-ended ds cDNAs. After digestion with Not I, the ds cDNAs flanked by Eco RI-Not I restriction sites were ligated to Eco RI-Not I digested λgt 22 vector arms. The recombinant DNA was packaged and transduced into *E. coli* LE 392 and Y1090. The cloning strategy is outlined in FIG. **2**.

Screening of λgt 22 cDNA Library

The λgt 22 cDNA library was screened, using as probe, *T. tridentatus* Factor C cDNA (pFC 53, also referred to as λFC 53), which is a partial fragment of the Factor C cDNA (Muta et al., 1991. *J. Biol. Chem.* 266, 6554–6561).

A. Primary Screening

1. Plating of Amoebocyte cDNA Library and Blotting to Nylon Membranes

Plating bacteria (*E. coli* LE392) were propagated in LB medium supplemented with 100mM $MgSO_4$ and 0.2% maltose (w/v). Cultures were grown to an OD_{600} of 0.6. In order to obtain approximately 250,000 pfu of the amoebocyte cDNA library on a Bioassay plate, (245×245 mm, Nunc), 14 μl of the undiluted phage stock containing 1.88×10^7 pfu/ml were mixed in a 50 ml Nunc centrifuge tube with 2 ml of plating bacteria. The mixture was then incubated at 37° C. for 20 min. to allow the phage particles to adsorb onto the host cells. Maltose is essential for bacterial expression of the lam b gene, which codes for the phage receptors, while magnesium ions facilitate the adsorption of phage particles to these receptors at 37° C. Thirty ml of molten top agarose were then added, and this was immediately mixed and poured onto LB bottom agar supplemented with 1 mM $MgSO_4$. The plate was incubated at 37° C. until the plaques were large enough to contain sufficient DNA for detection, but at the same time were not confluent with each other to allow easy plaque purification. A duplicate plate was similarly prepared to give a screening base of about 500,000 clones. Once the expedient phage size and confluency were achieved, incubation was terminated and the plates were transferred to 4° C. for 1 h. This allows the plaques to absorb enough moisture for them to stick properly to the filters during the blotting step.

A 1.2 μm nylon membrane (22×22 cm, Pall Biodyne A) was randomly labelled with a ballpoint pen on the four sides of one face and then applied face down (ink side up) on each of the two plates. The bottoms of each plate were then accurately marked corresponding to those found on the filters to record the orientation of the filters on the master plate. The filters were left for 2 min. to allow the phage particles to adsorb onto them. The filters were then peeled off slowly from one corner using a pair of forceps with flat ends, and subsequently air dried face up for at least 10 min. A replica filter was then applied onto the same plate, but this was allowed to stick for 10 min. before peeling off. Similarly, the orientation of the replica filter on the master plate was recorded, but the marks were fixed on different locations from the first filter. The replica filter serves to confirm positive clones as only truly hybridizing clones will show autoradiographic signals unambiguously at the same spot on two different filters.

2. Processing of Blotted Filters

After air drying, the phage DNA was denatured by incubating the filters in a solution of 0.2M NaOH/1.5M NaCl for 2 min. Subsequently, the phage DNA was neutralized with 0.4M Tris-HCl, pH 7.6/2×SSC for another 2 min.,

Seikagaku Corporation Exhibit 2004 - Page 271 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

and then saturated with 2×SSC for 2 min. The denatured phage DNA was then irreversibly immobilized onto the nylon membranes by UV irradiation at 312 nm for 2 min., followed by baking at 80° C. for 1 h.

3. Screening by Hybridization

After baking, membranes were prewashed in a solution of 50 mM Tris-HCl, pH 8.0, containing 1M NaCl 1 mM EDTA, and 0.1% SDS for 30 min. at room temperature to remove any bacterial debris still adhering to the filters. Filters were then prehybridized and hybridized as follows:

Prehybridization was carried out for 6 h at 42° C. in 50% formamide (v/v, Merck), 5×SSC (0.3M NaCl, 0.3M NaCl 0.3M Trisodium citrate), 5× Denhardt's solution (Maniatis et al. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, p. 327), 50 mM phosphate buffer, pH 6.5, and 0.1% SDS. Boiled calf thymus DNA (100 μg/ml, Sigma) was used as carrier DNA. Hybridization was carried out overnight at 42° C. in 25 ml hybridization buffer (50% formamide, 5×SSC, 1× Denhardt's solution, 20 mM phosphate buffer, pH 6.5, 50 μg/ml calf thymus DNA, 0.1% SDS) using ^{32}P-labelled Factor C cDNA of *T. tridentatus* (pFC53) as the probe. Salt and temperature conditions equivalent to the hybridization conditions employed can be calculated from the following equation (Sambrook et al. (1989) *Molecular Cloning, A Laboratory Manual*, Second edition, Cold Spring Harbor Laboratory Press, pp.9.50–9.51):

$$T_m = 81.5° C. - 16.6(\log_{10}[Na^+]) + 0.41(\%G+C) - 0.63(\%\text{formamide}) - (600/l),$$

where l=the length of the hybrid in base pairs.

The membrane was washed 3× at low stringency with 1×SSC, 0.1% SDS washing solution at room temperature, followed by washing at high stringency with 0.1×SSC, 0.1% SDS at 42° C. for 30 min. and 2× with 0.1×SSC/0.1% SDS for 15 min. each at 42° C., or until low background radioactivity was attained. Membranes were autoradiographed for 5 days with a Hyperfilm MP at −70° C. in the presence of an intensifying screen (Kodak X-Omatic or Dupont Lightning Plus). The film was developed using a Kodak X-Omat MP4 developer.

The Factor C cDNA probe equivalent to 200 ng was radioactively labelled using the Multiprime Labelling system.

B. Secondary and Tertiary Screening

Positive clones which were detected on the autoradiograms of both the main filter and its replica were traced back to their exact locations on the master plate. As the plaques were quite close to each other, a plug of agarose approximately 1 mm in diameter containing a number of plaques located within the most proximate area which gave the postive signal was picked from the plate using a sterile 100 μl Eppendorf tip, the end of which was snipped off. Plugs were resuspended in 1 ml of suspension medium (SM: 50 mM Tris-HCl, pH 7.5 containing 100 mM NaCl 8 mM $MgSO_4$ and 0.01% gelatin, (w/v)), containing a drop of chloroform for at least 1 h at room temperature to enable the phages to elute into the buffer solution.

An aliquot of the SM buffer containing eluted phages was then mixed with 200 μl of plating bacteria and plated a second time on LB plates (90 mm diameter) as described above, but such that only about 50 to 100 widely spaced plaques were formed. These were similarly plaque lifted onto nylon membranes, denatured, hybridized, washed, and exposed as described under "Primary Screening." A well-isolated plaque in the secondary screening was similarly resuspended in SM buffer, and a tertiary screening was performed. Tertiary screening confirms the purity of the secondary plaque by demonstrating whether all the plaques generated gave strong positive signals after autoradiography. A single plaque from the tertiary screening was then used to propagate and purify the presumptive clones for subsequent characterization.

As a rare copy gene, *C. rotundicauda* Factor C cDNA represented only 0.03% of the amoebocyte cDNA library. A total of 40 presumptive clones were isolated, of which 16 were further studied. The *C. rotundicauda* Factor C recombinants (λCrFC) were purified, and the sizes of their inserts were determined after digestion with Eco RI and Not I. There is one Eco RI site within these inserts. The CrFC cDNA partial fragments flanked by Eco RI-Eco RI (EE) and Eco RI-Not I (EN) were further subcloned into pGEM 11 Zf(+) and transformed into *E. coli* JM 109 and DH5α. These subclones, referred to as pEE and pEN, respectively, were characterized by restriction mapping and DNA sequencing.

EXAMPLE 2

Comparison of the Restriction Maps of CrFC and TtFC cDNAs

The orientations of the 16 λCrFC clones were determined, and subsequently, their initial restriction maps were analyzed for Eco RI and Not I sites (FIG. **3**). The complete restriction maps of the two longest clones of *C. rotundicauda* Factor C cDNAs, λCrFC 21 and λCrFC 26, are shown in FIG. **4**, in comparison with the maps of the two overlapping clones of *T. tridentatus* Factor C, TtFC (λFC 53 and λFC 41) (Muta et al., 1991. *J. Biol. Chem.* 266, 6554–6561). λCrFC 26 and λCrFC 21 will henceforth be referred to as CrFC 26 and CrFC 21, respectively. CrFC 26 has a longer 5' terminal end compared to the *T. tridentatus* FC 53. Based on the restriction sites of the λFC 53 and λFC 41 clones, the cDNAs from both species have the same locations for restriction sites such as Sty I, Sal I, Sac I, Hind III and Bam HI. However, notable differences were also observed between the cDNAs: CrFC 26 has two Bam HI sites compared to only one in λFC 53. On the other hand, the former has only two Hind III sites compared to three in the latter. The loss of one Hind III site in CrFC 21 and 26 was due to a base substitution (C to T) at the 1700 and 2443 bp positions, respectively.

λFC 41, on the other hand, encompasses the 3' region of the TtFC. It begins from the Sac I/Sst I site up to the 3' end, and was found to contain the internal Eco RI site relative to that of CrFC 26. The other common site between λFC 41 and CrFC 21 and 26 is the Pst I site proximal to the 3' end, but the second Pst I site near the 5' end of λFC 41 was absent in CrFC 21 and CrFC 26. The Bam HI and Hind III sites in λTtFC 41 were absent in CrFC 21 and CrFC 26.

EXAMPLE 3

The DNA Sequences and Derived Amino Acid Sequences of CrFC 26 and CrFC 21

The DNA sequences of CrFC 26 and CrFC 21 were determined on both strands by the Sanger dideoxy sequencing method. FIGS. **5***a* and 5b show the sequencing strategies for CrFC 26 and 21, respectively. To counter check the sequences obtained, the 5' ends of other CrFC subclones (pCrFC 69, 1, 16 and 35) were also sequenced using the T_7 promoter as primer binding site. The complete cDNA of CrFC 26 was found to be 4182 bp (FIG. **6**). This is consistent with the Northern analysis (Reed and Mann, 1985. *Nucl. Acid Res.* 13, 7207–7221) of the amoebocyte total RNA with homologous EE insert, which showed only one hybridizing band at approximately 4 kNt (FIG. **7***a*). CrFC 26 cDNA

Seikagaku Corporation Exhibit 2004 - Page 272 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

includes 568 nucleotides of 5' untranslated sequence containing seven ATGs before the real initiation site, an ORF of 3249 nucleotides coding for a protein of 1083 amino acids, a stop codon, and 365 nucleotides of 3' untranslated sequences. At the 3' end of the cDNA, the canonical hexanucleotide sequence, AATAAA, is present 19 nucleotides upstream from the polyadenylation site (at nucleotide position 4142). There are three other AATAAA sequences, found at nucleotide positions 183, 239 and 2474, respectively. The ORF codes for a signal peptide of 24 amino acids and a Factor C zymogen of 1059 residues, which, prior to N-glycosylation, has a calculated molecular weight of 120, 244 daltons. This is close to the estimated molecular weight of single- and double-chain Factor C enzymes purified from amoebocytes (Ding et al., 1993. *Biochim. et Biophys. Acta.* 1202, 149–156). The unique proteolytic site due to endotoxin-activation of the Factor C enzyme is found between phe-ileu, thus indicating the integrity of the cloned Factor C cDNA, and its potential application in endotoxin detection.

The CrFC 21 cDNA sequence and its derived amino acid sequence are shown in FIG. **8**.

EXAMPLE 4

Variant Forms of the CrFC Gene

The Factor C cDNA sequence indicates the existence of at least two types of Factor C mRNA in C. rotundicauda. The N-terminal sequences of CrFC 26 and CrFC 21 differed significantly. Comparison of the sequences between *C. rotundicauda* Factor C clones CrFC 21 and CrFC 26 with that of *T. tridentatus* (TtFC 53) shows that CrFC 21 and TtFC 53 share precisely similar homology at the 5' end. Yet, the most notable difference in the DNA sequences of Factor C from the two species of horseshoe crabs is the existence of an extra 716 nucleotides at the 5' end of the CrFC 26 (FIG. **9**). Thus, CrFC 21 was not merely a truncated species of CrFC 26. Instead, it was derived from a totally distinct species of mRNA that may be exactly identical to the one that gave rise to the *T. tridentatus* Factor C cDNA. To unequivocally demonstrate the authenticity of the 5' noncoding region of CrFC 26, a further Northern blot analysis was performed, but using only the first 369 bp of the 5' untranslated region of CrFC 26 (fragment flanked by Eco RI and Nde I) as probe. The result in FIG. **7***b* clearly demonstrates that the entire 5' end of CrFC 26 truly belongs to this particular species of Factor C mRNA. Thus, the existence of the two types of *C. rotundicauda* Factor C cDNA could be attributable to differential splicing of the initial primary transcript around its 5' terminal.

EXAMPLE 5

Computational Analysis of the 5' Noncoding Region of λCrFC 21 and 26

The unusually long 5' untranslated sequences occur on unusually interesting mRNAs such as epidermal growth factor, EGF (Scott et al., 1983. *Science* 221, 236–240), oncogenes (Watt et al., 1983. *Nature* 303, 725–728) and heat shock proteins (Ingolia et al., 1981. *Nucl. Acids Res.* 9, 1627–1642). This invites speculation that long structures of the 5' noncoding region participate in the regulated expression of these genes, including that of Factor C. However, long 5' noncoding sequences in some mRNAs may have a deleterious effect (Kozak, 1983. *Microbiol. Rev.* 47, 1–45), especially if they contain secondary structures which may deflect ribosomes from the authentic initiation site that lies further downstream. This was true for the Semliki Forest virus (Lehtovaara et al., 1982. *J. Mol. Biol.* 156, 731–748) and the recombinant fish antifreeze protein III gene (Li et al., 1991. *Protein Eng.* 4, 995–1002). This could also apply to CrFC 26 since there are seven false ATG sites (underlined in FIG. **6**) lying upstream from the authentic ATG start site found at nucleotide position 569. All these false start sites were followed shortly by in-frame stop sites downstream. This configuration agrees with the ribosome scanning-reinitiation model of translation in eukaryotic genes (Kozak, 1986. Cell 47, 481–483). This is probably one reason why there exists another form of Factor C mRNA having a much shorter 5' noncoding sequence as in CrFC 21, where it may be translated more efficiently.

FIGS. **10***a* and **10***b* show the secondary structure predicted for the 5' ends of the mRNA of CrFC 21 and CrFC 26, respectively. CrFC 21 exhibits a well-exposed AUG codon in a loop. In contrast, CrFC 26, with its 7 false start codons, shows numerous hairpin stems and loops, and its real start codon appears less accessible to ribosome binding.

From computational analysis of the secondary structures of CrFC 21 and 26, it is clear that both clones could be further manipulated at their 5' ends to give appropriate Factor C cDNAs for subcloning into *Pichia pastoris* vector for overexpression of the recombinant Factor C enzyme. The recombinant Factor C could be used in a colorimetric assay for endotoxin.

EXAMPLE 6

Homology Between the DNA Sequences of CrFC and TtFC

In comparing the cDNAs of CrFC 26 and 21 with that of TtFC (FIG. **11**), the *C. rotundicauda* gene was found to be longer by 716 nucleotides at the 5' end, and also has an extra 64 amino acids upstream from the starting met. The start of homology of both the DNA and amino acid sequences between the two species was found at nucleotide positions 785 for CrFC 26 and 42 for CrFC 21, as compared to position 69 for TtFC. The homologous region extends through to the poly A tail of both genes, giving 97.7% homology between CrFC 26 and TtFC. Most of the dissimilarities, however, were found in the 3' untranslated region, where a lower percentage homology of 86% was observed.

A majority of the differences between the CrFC and TtFC sequence found within the ORF were due to base substitutions. In fact, these substitutions explain the perceived differences in the restriction sites earlier obtained by restriction mapping (see FIG. **4**), for instance, the absence of the Bam HI (GGATCC) site in TtFC, which in CrFC 21 and 26 is located 807 and 1550 bp, respectively, from their 5' ends. This was due to the substitution of the first G by T in TtFC. The converse is shown with the Bam HI site at bp 2431 of TtFC, in which the second C is substituted by a T in the CrFC cDNAs. Similarly, the Pst I site (CTGCAG) of TtFC (at 1688 bp from its 5' end) was not found in CrFC because the first C was substituted by a T. Again, base substitution of C by T was responsible for the loss of the third Hind III site (AAGCTT) in CrFC, which in TtFC was located 1725 bp from the 5' end. Base substitution also occurred at 2361 and 3104 bp of CrFC 21 and 26, respectively, giving the Xba I (TCTAGA) site, where the first A is substituted by a C in TtFC. Similar to TtFC, there are six potential N-glycosylation sites within the CrFC sequence Asn-Xaa-Ser/Thr.

EXAMPLE 7

Structural Domains and Homology Comparison of CrFC With Other Serine Proteases

Seikagaku Corporation Exhibit 2004 - Page 273 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

The secondary folding structures of several blood clotting factors in the human coagulation system are usually identified by common domains derived from homologous regions in the DNA sequence (Furie, B. and Furie, B.C. 1988. *Cell* 53:505–518). Typically, all these factors have a signal peptide at their N-terminal, and a catalytic domain at the C-terminal, which for serine proteases like *C. rotundicauda* Factor C, also contain the catalytic triad, His^{809}-Asp^{865}-Ser^{966} for CrFC 21 and His^{873}-Asp^{929}-Ser^{1030} for CrFC 26 (see FIG. **11**). A hydropathy profile based on the formula of Kyte and Doolittle (1982, *J. Mol Biol.* 157, 105–132) of the first 59 amino acids of CrFC 26 showed the first 24 residues to be the putative signal peptide rich in hydrophobic amino acids (FIG. **12**). CrFC 26 has five short consensus repeats (residues 206–259, 263–318, 324–385, 640–698 and 758–812), one lectin-type domain (500–632), one EGF-like domain (167–200), a cys-rich region (90–165), and one pro-rich region (732–754). The serine protease domain is found at positions 827–1083). No propeptide sequence, kringle domains (Park, C. H. and Tulinsky, A. 1986. *Biochemistry* 25, 3977–3982), amino acid stacks, or finger domains were found.

By scanning the EMBL gene data bank, a homology search for the serine protease catalytic domain of CrFC revealed that it is structurally closest to *T. tridentatus* Factor C, with 97.7% homology in 777 overlapping nucleotides (Table I).

TABLE I

Percentage homology comparison of the catalytic domain of *C. rotundicauda* Factor C with other serine proteases based on entries in the EMBL data bank.

Gene	% homology	bp overlap
Tachypleus tridentatus Factor C	97.7	777
Onchorhynchus mykiss thrombin	54.4	709
Rattus norvegicus thrombin	52.2	624
Rat prothrombin	52.2	624
Acipenser transmontanus thrombin	52.8	623
Oryctolagus cuniculus thrombin	57.3	595
Mouse prothrombin	52.7	509
Human blood coagulation Factor VII	55.7	420
Guinea pig Factor IX	57.4	296
Human liver hepsin	59.0	278

From the overall comparison with serine proteases from other species, the CrFC is structurally closest to prothrombin and thrombin. This finding agrees well with the preference of Factor C for hydrolysis of synthetic substrates of thrombin. Surprisingly, a human liver hepsin (Leytus et al., 1986. *Biochemistry* 25, 5098–5102) exhibited a slightly higher level of homology with CrFC than other blood coagulation factors. Although the function of hepsin is still unknown, this serine protease, which lacks a typical signal peptide but instead has a transmembrane domain, was nevertheless shown to have characteristics typical of trypsin. Based on the primary structure of CrFC, it was also found to be catalytically similar to trypsin-type serine protease, although structurally closer to thrombin. This may be explained by the Asp^{1024} (as opposed to Ser) in CrFC 26, analogous to Asp^{189} in trypsin (Hartley, B. S. 1970. *Phil. Trans. R. Soc. London*, 257, 77–87), which strongly suggests that CrFC possesses a substrate specificity similar to that of trypsin.

EXAMPLE 8

Subcloning of CrFC cDNAs in Expression Vectors

The λCrFC cDNAs were further manipulated and subcloned into various plasmid vectors. Firstly, for easier manipulations of the clones, the complete Factor C cDNAs that were originally cloned in λgt 22 (clones 21 (3448 bp) and 26 (4182 bp), each being flanked by multicloning sites: Eco RI and Not I) were recloned into plasmid, pGEM11Zf (+) to give pGEM11Zf(+)/CrFC 21 and pGEM11Zf(+)/CrFC 26, respectively. The recloning strategy is outlined in FIG. **13**. Secondly, restriction digestions of the Factor C insert in recombinant clone λCrFC 21 with Sal I and Not I yielded a 2.4 kb fragment which was then subcloned into expression vector pGEMEX-1 (FIG. **14**). Thirdly, digestion of λCrFC 21 with Bgl II and Eco RV resulted in a 3388 bp Factor C insert which was subcloned into pET 3b expression vector pAR 3039 that was previously linearized using Bam HI and Eco RV (FIG. **15**). Fourthly, since the gene may not be apparently expressed at detectable levels in vivo in the *E. coli* host (possibly due to activation of the gene product by the Gram negative host bacterial endotoxin, resulting in the active Factor C being toxic to the host), the CrFC 26 was also inserted into yeast shuttle secretory (YEpsec1) and non-secretory (pEMBLyex4) expression vectors (FIG. 16). The CrFC 21 and CrFC 26 were further excised from their pGEM11Zf(+) vector and subcloned into vectors capable of transforming them into the methylotrophic yeast, *Pichia pastoris* (FIG. **17**). The superiority of the *P. pastoris* system lies in the AOX1 promoter-driven expression of the Pichia vectors, which would result in high levels of expression of the CrFC cDNAs. The rationale behind the use of the yeast system is that (a) the gene under study is from a higher eukaryote whose product needs to be post-translationally modified, and (b) advantage will be taken of the absence of endotoxin in the yeast host to obtain the inactive Factor C proenzyme.

cDNAs And Proteins For Factor C

Each of the nucleic acid sequences and polypeptides disclosed herein, or their biologically functional equivalents, can be used in accordance with the present invention. The term "biologically functional equivalents," as used herein, denotes nucleic acid sequences and polypeptides exhibiting the same or similar biological activity as the particular nucleic acid sequences and polypeptides described herein.

For example, the nucleic acid sequences described herein can be altered by base substitutions, additions, or deletions to produce biologically functionally equivalent nucleic acids that encode proteins exhibiting Factor C enzymatic activity in endotoxin assays. In addition, due to the degeneracy of the genetic code, other DNA sequences that encode substantially the same amino acid sequences as described herein exhibiting Factor C enzymatic activity in endotoxin assays may be used in the practice of the present invention. These include, but are not limited to, nucleotide sequences comprising all or portions of the Factor C cDNAs described herein which are altered by the substitution of different codons that encode a physiologically functionally equivalent amino acid residue within the sequence, thus producing a silent change. Similarly, the Factor C proteins, or derivatives thereof, of the present invention include, but are not limited to, those containing all of the amino acid sequences substantially as described herein, including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence, resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted with another amino acid of similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine,

leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

The variants of Factor C cDNAs and proteins contemplated herein should possess more than 75% homology, preferably more than 85% homology, and most preferably more than 95% homology, to the naturally occurring Factor C cDNAs and proteins discussed herein. To determine this homology, two proteins are aligned so as to obtain a maximum match using gaps and inserts. Homology is determined as the product of the number of matched amino acids divided by the number of total amino acids plus gaps and inserts, multiplied by 100.

Also included within the scope of the present invention are Factor C fragments or derivatives thereof which are differentially modified during or after translation, e.g., by glycosylation, proteolytic cleavage, etc.

In addition, the recombinant Factor C-encoding nucleic acid sequences of the present invention may be engineered so as to modify processing or expression of Factor C. For example, and not by way of limitation, a signal sequence may be inserted upstream of Factor C encoding sequences to permit secretion of Factor C, and thereby facilitate harvesting or bioavailability.

Additionally, a given Factor C isoform or mutein can be mutated in vitro or in vivo, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further in vitro modification. Any technique for mutagenesis known in the art can be used, including, but not limited to, in vitro site-directed mutagenesis (Hutchinson et al. (1978) *J. Biol. Chem.* 253:6551), use of TAB® linkers (Pharmacia), etc.

Expression Vectors For Factor C

The vectors contemplated for use in the present invention include those into which a DNA sequence as discussed herein can be inserted, along with any necessary operational elements. Such vectors can then be subsequently transferred into a host cell and replicated therein. Preferred vectors are those whose restriction sites have been well documented and which contain the operational elements preferred or required for transcription of the DNA sequence.

Certain embodiments of the present invention employ vectors which would contain one or more of the DNA sequences described herein. It is preferred that all of these vectors have some or all of the following characteristics: (1) possesses a minimal number of host-organism sequences; (2) be stably maintained and propagated in the desired host; (3) be capable of being present in high copy number in the desired host; (4) possess a regulatable promoter positioned so as to promote transcription of the gene of interest; (5) have at least one marker DNA sequence coding for a selectable trait present on a portion of the plasmid separate from that where the DNA sequence will be inserted; and (6) contain a DNA sequence capable of terminating transcription.

The cloning vectors capable of expressing the DNA sequences of the present invention contain various operational elements. These "operational elements" can include at least one promoter, at least one initiator codon, and at least one termination codon. These "operational elements" may also include one or more of the following: at least one operator, at least one leader sequence for proteins to be exported from intracellular space, at least one gene for a regulator protein, and any other DNA sequences necessary or preferred for appropriate transcription and subsequent translation of the cloned Factor C DNA.

Certain of these operational elements may be present in each of the preferred vectors of the present invention. It is contemplated that any additional operational elements which may be required may be identified and added to these vectors using methods known to those of ordinary skill in the art, such as those described by Sambrook et al. (1989) *Molecular Cloning, A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press.

Regulators

Regulators serve to prevent expression of the DNA sequence in the presence of certain environmental conditions and, in the presence of other environmental conditions, will allow transcription and subsequent expression of the protein coded for by the Factor C DNA sequences. Regulatory segments can be inserted into the vector such that expression of the DNA sequence will not occur, or will occur to a greatly reduced extent. Expression of the desired protein is induced by addition of a substance to the environment capable of causing expression of the DNA sequence after the desired cell density has been reached.

Promoters

The expression vectors must contain promoters which can be used by the host cell for expression of its own proteins. Many promoters have been isolated and characterized, enabling one skilled in the art to use them for expression of the recombinant Factor C forms.

Transcription Terminators

The transcription terminators contemplated herein serve to stabilize the vector. Those sequences described by Rosenberg et al. (1979) *Ann. Rev. Genet.* 13:319–353 can be used in the present invention.

Non-translated Sequences

It may also be desirable to construct the 3' or 5' end of the coding region to allow incorporation of 3' or 5' non-translated sequences into the cDNA transcript. Included among these non-translated sequences are those which stabilize mRNA, as disclosed by Schmeissner et al. (1984) *J. Mol. Biol.* 176:39–53.

Leader Sequences and Translational Couplers

Additionally, DNA coding for an appropriate secretory leader (signal) sequence can be present at the 5' end of the DNA sequence, as set forth by Watson, M.E. in *Nucleic Acids Res.* 12:5145–5163, if the protein is to be excreted from the host cytoplasm. The DNA for the leader sequence must be in a position that allows the production of a fusion protein in which the leader sequence is immediately adjacent to and covalently joined to Factor C, i.e., there must be no transcription or translation signals between the two DNA coding sequences.

In some species of hosts, the presence of an appropriate leader sequence will allow transport of the completed protein into the periplasmic space. In the case of some Saccharomyces, the appropriate leader sequence will allow transport of the protein through the cell membrane and into the extracellular medium. In this situation, the protein may be purified from other extracellular proteins.

Translation Terminators

The translation terminators contemplated herein serve to stop the translation of mRNA. They may be either natural, as described by Kohli, *J., Mol. Gen. Genet.* 182:430–439, or synthetic, as described by Pettersson, R.F. (1983) Gene 24:15–27.

Selectable Markers

Additionally, the cloning vectors contemplated herein can contain a selectable marker, such as a drug resistance marker or other marker which causes expression of a selectable trait by the host cell.

Such a drug resistance or other selectable marker facilitates the selection of transformants. Additionally, the presence of such a selectable marker in the cloning vector may be of use in keeping contaminating microorganisms from multiplying in the culture medium. A pure culture of the transformed host cells would be obtained by culturing the cells under conditions which require the induced phenotype for survival. The operational elements discussed herein are routinely selected by those of ordinary skill in the art in light of prior literature, including Sambrook et al., discussed supra, and the teachings contained herein. General examples of these operational elements are set forth in B. Lewin (1983) *Genes*, Wiley & Sons, New York. Various examples of suitable operational elements may be found in the vectors discussed above, and may be gleaned via review of the publications discussing the basic characteristics of the aforementioned vectors.

Upon synthesis and isolation of all the necessary and desired component parts, the vector can be assembled by methods generally known to those of ordinary skill in the art. Assembly of such vectors is within the ordinary skill in the art, and, as such, is capable of being performed without undue experimentation.

Multiple copies of the DNA sequences of the present invention and their accompanying operational elements may be inserted into each vector. In such case, the host cell would produce greater amounts per vector of the desired form of Factor C. The number of multiple copies of the DNA sequence which may be inserted into the vector is limited only by the ability of the resultant vector, due to its size, to be transferred into and replicated and transcribed in an appropriate host cell.

Host Cells

Vectors suitable for use in various host cells are contemplated for use in the present invention. Such host cells include, for example, *E. coli*, yeasts such as *Saccharomyces cerevisiae* and *Pichia pastoris*, Baculovirus, and mammalian cells, including, for example, Chinese Hamster Ovary cells.

In the case of yeasts, useful promoters include Gal 1 and 10, Adh 1 and 11, and Pho 5. Transcription terminators can be chosen from among Cyc, Una, Alpha Factor, and Sac 2. Transcriptional start sites and leader peptides can be obtained from the invertase, acid phosphatase, and Alpha factor genes. Useful selection markers are Ura 3, Leu 2, His 3, and Tap 1.

In the case of expression in mammalian cells, the DNA encoding the present forms of Factor C should have a sequence efficient at binding ribosomes. Such a sequence is described by Kozak in *Nucl. Acids Res.* (1987) 15:8125–8132. The Factor C-encoding fragment can be inserted into an expression vector containing a transcriptional promoter and a transcriptional enhancer as described by Guarente in *Cell* (1988) 52:303–305 and Kadonaga et al. (1987) *Cell* 51:1079–1090. A regulatable promoter as in the Pharmacia plasmid pMSG can be used, if necessary or desired. The vector should also possess a complete polyadenylation signal as described by Ausubel et al. (1987) in *Current Protocols in Molecular Biology*, Wiley, so that mRNA transcribed from the vector is properly processed.

In order to select a stable cell line that produces Factor C as described herein, the expression vector can carry the gene for a selectable marker such as a drug resistance marker or a complementary gene for a deficient cell line, such as a dihydrofolate reductase (dhfr) gene for transforming a dhfr$^-$ cell line, as described by Ausubel et al., supra. Alternatively, a separate plasmid carrying the selectable marker can be cotransformed along with the expression vector.

Vectors for mammalian cells can be introduced therein by several techniques, including calcium phosphate:DNA coprecipitation, electroporation, or protoplast fusion. Coprecipitation with calcium phosphate as described by Ausubel et al., supra, is the preferred method.

By way of example, vectors and host cells contemplated for the heterologous expression of the Factor C cDNAs disclosed herein include the use of the T7 system in pGEMEX1, pET3b(pAR3039), and pGEM11Zf(+), and the GST system in pGEX1,2,3. The host cells can be chosen from among JM109, JM109(DE3), DH5α, HMS174(DE3) plysS, and λCE6(phage).

For expression in *Pichia pastoris*, the host strain can be GS115, and useful vectors are pHIL D2, pPIC 9, and pHIL S1.

For expression in the Baculovirus system, host cells can be *Spodoptera frugiperda* SF9 and SF21; the vectors can be pBlue Bac His, A, B, C; pBlue Bac III; pVL1392; pVL 1393; and pAC360. The host for initial cloning can be *E. coli* DH5α, JM109, or TOP 10F.

For expression in mammalian cells, the vector for transfection can be pCDNAI, wherein all initial cloning is in *E. coli* MC1061/PE3. Mammalian host cells can be African green monkey derived COS1 or COS7 cells which express the SV40 large T antigen. The mouse fibroblast cell line NIH3T3, which also expresses the SV40 large T antigen, can also be used.

EXAMPLE 9

Expression of Factor C cDNA Constructs In Vitro

The CrFC cDNAs subcloned into pGEM11Zf(+), pGEMEX-1, and pET 3b were subjected to in vitro transcription and translation using the Promega transcription and translation system (TnT T_1 coupled rabbit reticulocyte lysate). The expression of the cDNA inserts was driven by T7 promoter in their respective vectors. Using 2 μg each of the recombinant DNA construct, variable rates of transcription and translation were observed, giving the expected sizes of either fusion or non-fusion Factor C gene products (FIG. 18). These results therefore demonstrate that the CrFC cDNAs were subcloned in frame with their T_7 promoters. The pGEM11Zf(+)/CrFC 26 full-length cDNA was not expressed, possibly due to its high number of false start codons (see FIG. 6).

The deletion subclones of CrFC 26, viz., pGEM11Zf(+) /CrFC 26–6a and –9a, which lack the 5' untranslated region of CrFC 26, were better expressed, yielding faint bands of Factor C proteins. Deletion subclones 6a and 9a were created by carrying out 5' –3' Exo III nuclease mutagenesis on the Eco RI fragment of CrFC 26 cDNA which had been cloned into the unique Eco RI site of pGEM11Zf(+). In order to perform the deletion mutagenesis, the recombinant clones was first digested with Not I, and the recessed ends Klenow-filled with α-phosphorothioate nucleotides to create a protected site resistant to Exo III nuclease digestion. The clone was then subsequently digested with Xba I to create a sensitive site from which Exo III nuclease digestion could initiate. The deletion mutagenesis was carried out at 30° C., and aliquots of the reaction mixture were removed at 30-second intervals to produce a series of deletion mutants. Subclones containing deletions of appropriate sizes were sequenced using Sanger's dideoxy method. Deletion sub-

Seikagaku Corporation Exhibit 2004 - Page 276 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

clones 6a and 9a were selected for expression analyses because they contain complete deletions of the 5' untranslated DNA sequence, and also partial deletions of the DNA sequence encoding the putative leader peptide. Deletion subclone 9a starts at nucleotide position 721, while subclone 6a starts at nucleotide position 762. Complete deletion of the 5' untranslated DNA sequence from both subclones is expected to release their expression from the control of any translational signals which may be encoded in the 5' untranslated sequence. Manipulation of the CrFC 26 sequence to eliminate some or all of the false start codons is therefore expected to produce deletion subclones which can be successfully expressed, producing active Factor C.

On the other hand, the pGEM11Zf(+)/CrFC 21 full-length cDNA construct is more efficiently expressed than its CrFC 26 counterparts because CrFC 21 has the true ATG and Kozak consensus sequence belonging to the CrFC 21 cDNA insert itself, which makes it more complete and translatable. The 6a and 9a constructs were not driven by the CrFC 26's own true ATG. Rather, The ATG in both cases was present in the multicloning site of pGEM11Zf(+). However, in pGEMEX-1 vector, CrFC 21 (pGEMEX-1/CrFC 21 construct) yielded the highest level of incorporation of ^{35}S-cys and a correspondingly more intense band of translated Factor C gene product. This may be due to the fact that the fusion 260 amino acids code for the T_7 gene10 capsid protein and is thus best transcribed under its own T_7 promoter using the compatible T_7 RNA polymerase. This efficient expression of CrFC 21 was also observed in pET 3b vector.

These results indicate that the present CrFC cDNA recombinant clones can be expressed in heterologous systems, depending on the manner of insertion of the cDNAs into appropriate vectors.

Purification of Factor C Proenzyme

Attempts to purify Factor C proenzyme from amoebocyte lysates have often been hampered by the ubiquitous endotoxin which activates the zymogen into an active serine protease enzyme. This results in poor yields as well as considerable loss of enzymatic activity. In order to improve the purification procedure, dimethylsulfoxide, Me_2SO, and chelating agents, such as EDTA, have been incorporated in the buffer solution to prevent the premature activation of the cascade reaction during purification. This has led to a simple and unique procedure for concomitantly purifying two isoforms of Factor C precursor enzymes, with a corresponding amelioration of their total yields and specific activities. The data presented infra show that both the single- and double-chain forms of the enzyme have endotoxin receptors to which endotoxin binds to activate their catalytic sites. Data presented infra also show that these endotoxin-binding sites in Factor C are competitively but reversibly occupied by Me_2SO in the range of from about 5% to about 30% when the latter is added during purification.

EXAMPLE 10

Preparation of Amoebocyte Lysate from *C. rotundicauda*

Horseshoe crabs were bled by cardiac puncture as previously described by Jorgensen and Smith, 1973. *Appl. Microbiol.* 26, 43–48. Blood was collected into a solution of 0.125% N-ethylmaleimide/3% NaCl (w/v) and centrifuged at 150× g for 30 min at 26° C. Amoebocytes were washed with 3% NaCl and lysed overnight with pyrogen-free water at 4° C. The lysate was lyophilized by freeze-drying.

EXAMPLE 11

Purification of Single- and Double-Chain Factor C Proenzymes

The first step in the purification of Factor C is as previously reported (Navas et al., 1990. *Biochem. Intl.* 21, 805–813), except for the addition of Me_2SO, preferably in an amount of about 5%, v/v, and a chelating agent, such as EDTA, preferably in an amount of about 1 mM, in the gel filtration buffer. Freeze-dried Carcinoscorpius amoebocyte lysate (CAL) containing 326 mg total protein was reconstituted in 4 ml pyrogen-free water before fractionation on Sepharose CL-6B (2.6×97 cm) which was previously equilibrated with 0.05M Tris-HCl, pH 8.0, containing 0.154M NaCl 5% Me_2SO (v/v), and 1 mM Na_2EDTA. The flow rate was maintained at 20 ml/h. The fractions for each peak were pooled separately and lyophilized. The freeze-dried fractions were reconstituted in pyrogen-free water and desalted through Sephadex G-25 (1×6 cm) packed in a BioRad 10-DG column using pyrogen-free water as the mobile phase. Desalting effected the removal of Me_2SO together with all the other salts from the enzyme fractions. Me_2SO apparently hinders the binding of Factor C to the subsequent affinity column. The desalted enzyme fractions were separately subjected to affinity chromatography through a heparin-Sepharose CL-6B column (1×11 cm) at 25° C. using the Pharmacia FPLC system. The mobile phase contained 0.02M Tris-HCl, pH 8.0, with 1 mM Na_2EDTA. Protein (A_{280nm}) was eluted with a linear gradient of 0–0.5M NaCl at a flow rate of 0.75 ml/min and immediately stored at 4° C. This step essentially purified single-chain Factor C to apparent homogeneity, while double-chain Factor C was still co-eluted with proclotting enzyme. The fractions containing the double-chain Factor C were pooled and concentrated to a volume of 0.2 ml, and further purified by gel filtration in a Pharmacia FPLC Superose 12 column (HR 10/16) using 0.02M Tris-HCl, pH 8.0, containing 1 mM Na_2EDTA as the mobile phase.

The chromatograms from gel filtration (Sepharose CL-6B) of Factor C in the absence and presence of 5% Me_2SO and EDTA are compared in FIGS. **19***a* and **19***b*. In the absence of Me_2SO and EDTA, both Factor C and proclotting enzyme co-eluted in the first peak with an apparent native molecular mass of 495 kDa (FIG. **19***a*). This confirmed earlier observations (Wright and Jong, 1986. *J. Exp. Med.* 164, 1876–1888) that the coagulation enzymes, being glycoprotein in nature, tend to aggregate in the absence of a denaturing agent. The suspected single-chain Factor C (band A, FIG. **20***a*) was seen to be coincident with the A_{280nm} elution profile of the first peak in FIG. **19***a*. However, the appearance of a band which corresponds to the heavy chain of the clotting enzyme (band B in FIG. **20***a* and band A in FIG. **21**) implied that autoactivation (Nakamura et al., 1985. *J. Biochem.* 97, 1561–1574) occurred during the chromatography, thus suggesting that Factor C could have been transformed to its active form. Morever, it was also found that during some routine preparations of Factor C, coagulogen (band B, FIG. **21**) was prematurely transformed to coagulin (band C, FIG. **21**).

To circumvent this problem, the same chromatographic run was performed in the presence of Me_2SO and EDTA as shown in FIG. **19***b*. Under this condition, autoactivation was precluded and there was a substantial difference in the elution profile. A small protein peak, where both Factor C and proclotting enzyme were found to be concentrated, appeared between the first two peaks of the previous run. By SDS-PAGE analysis, single-chain Factor C (band A, FIG.

Seikagaku Corporation Exhibit 2004 - Page 277 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

20*b*) was also shown to be more well-separated from other contaminating proteins. However, in the presence of Me_2SO and EDTA, this band of single-chain Factor C did not seem to exhibit any activity. Instead, another form of Factor C was suspected to be responsible for such activity coincident with fractions 30–38 (bands C and D, FIG. 20*b*). This was attributable to the heavy- and light-chain, respectively, of the double-chain form of Factor C. Both peaks 2 and 3 in FIG. 19*b* were separately pooled, concentrated, and subjected to desalting through Sephadex G-25 using pyrogen-free water as eluent. With the removal of Me_2SO, the pooled first peak fractions regained a considerable amount of Factor C activity, contributing 30% of the total. This can be explained by earlier kinetic studies (Navas et al., 1990. *Biochem. Intl.* 21, 805–813) which showed that Me_2SO reversibly inactivates single-chain Factor C by binding to its endotoxin receptor site. Furthermore, single-chain Factor C was found to be more enriched at this stage, as compared to that in the previous purification. FIG. 22 shows the elution profile of single-chain Factor C. By this method, single-chain Factor C was purified 563-fold, to apparent homogeneity, giving a yield of 41% (Table II). By the same method, however, the double-chain Factor C still co-eluted with proclotting enzyme. Thus, further purification performed by gel filtration on Superose 12 using FPLC (FIG. 23) finally isolated the double-chain Factor C, giving 1520-fold purification (Table III).

were found to possess properties typical of serine proteases. In particular, both were irreversibly inhibited by small amounts of DFP and PMSF. However, both forms of Factor C reacted differently in the presence of Me_2SO. Single-chain Factor C was highly susceptible to Me_2SO, being almost completely inhibited by 5% Me_2SO, whereas double-chain Factor C activity was completely inhibited only at 30% Me_2SO (FIG. 25). Being more susceptible to Me_2SO, single-chain Factor C did not exhibit any apparent activity during the initial gel filtration step, where the Factor C assay was performed in the presence of 5% Me_2SO. This differential sensitivity of the two forms of Factor C to Me_2SO was instrumental in distinguishing the assay of one form from the other, and in the final purification simultaneously, of the two forms of Factor C. Thus, Me_2SO4 in the range of from about 5% to about 30%, v/v, can be used to inhibit Factor C.

EXAMPLE 12

Endotoxin Binding Assay

To test the presence of endotoxin (LPS) binding sites on Factor C, 10 μg aliquots of endotoxin were electrophoresed in duplicate in 15% polyacrylamide gels containing SDS. One set was silver stained according to the method of Tsai and Frasch, 1982. Anal. Biochem. 119, 115–119, while the other was electroblotted onto an Immobilon PVDF membrane (Millipore) using the same procedure as described above. The electroblotted membrane containing endotoxin

TABLE II

Purification of single-chain Factor C

Factor C activity from the Sepharose CL-6B column was determined after desalting the protein fractions through Sephadex G-25 in order to reactivate the enzyme by the removal of DMSO.

		Protein		Enzyme				
Purification step step	Volume (ml)	mg/ml	Total (mg)	U/ml	Total (U)	U/mg	Recovery (%)	Purification (-fold)
CAL	4	81.50	326.00	74.25	297	0.91	100	1
Sepharose CL-6B	40	0.04	1.60	5.82	233	155.33	78	171
Heparin Sepharose CL-6B	5	0.05	0.25	24.54	123	512.50	41	563

TABLE III

Purification of double-chain Factor C

		Protein		Enzyme				
Purification step	Volume (ml)	mg/ml	Total (mg)	U/ml	Total (U)	U/mg	Recovery (%)	Purification (-fold)
CAL	4.0	81.50	326.00	74.25	297	0.91	100	1
Sepharose CL-6B	35.5	0.08	2.84	7.21	256	90.14	86	99
Heparin Sepharose CL-6B	5.0	0.24	1.20	36.84	184	153.33	62	168
Superose 12	1.0	0.06	0.06	83.00	83	1383.33	28	1520

By non-reducing and reducing SDS-PAGE analysis, the single-chain Factor C remained as a single polypeptide with a calculated apparent molecular mass of 132 kDa (FIG. 24, lanes 2 and 4), while its double-chain form was found to be composed of a heavy (80 kDa) and a light chain (52 kDa) when resolved under reducing SDS-PAGE (FIG. 24, lane 5). The molecular masses reported here are slightly heavier than those reported for *T. tridentatus* (Nakamura et al., 1986. *Eur. J. Biochem.* 154, 511–521). Comparing the effective sizes of their component chains, it is apparent that the light chain of the enzyme from *C. rotundicauda* was heavier than that from *T. tridentatus*. Both forms of Factor C from *C. rotundicauda* was blocked by incubating in 50 mM Tris-HCl, pH 8.0, containing 0.2M NaCl (TBS) with 30 mg/ml BSA for 30 min at 37° C. The membrane was cut into strips (LPS strips), and each strip was separately incubated with slight agitation at 37° C. with 5 μg each of purified single-chain and double-chain Factor C fractions. The strips were then washed three times for 5 min each with TBS before incubation for 3 h at 37° C. with rabbit anti-CAL antiserum diluted 500-times in TBS containing 1 mg/ml BSA. A control experiment was carried out by incubating replicate LPS strips earlier treated with single- or double-chain Factor C with pre-immune rabbit serum instead of rabbit anti-CAL antiserum.

Seikagaku Corporation Exhibit 2004 - Page 278 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

Subsequently, the strips were washed with TBS, followed by incubation for 1 h at 37° C. with peroxidase-conjugated goat anti-rabbit IgG (Zymed) in TBS with 1 mg/ml BSA. After rinsing extensively, the strips were stained with 60 µl H_2O_2 and 60 mg chloro-1-napthol (Sigma) in 20% methanol (v/v).

FIG. **26** shows that LPS electroblotted to an Immobilon membrane was recognized by both the single-and double-chain Factor C. The single-chain Factor C showed higher capacity for binding to LPS as indicated by the stronger hybridization signal (FIG. **26**B, lane 1) compared to its double-chain counterpart. Since LPS and Me_2SO bind to the same receptor on Factor C, this observation is supported by the earlier finding of the higher susceptibility of single-chain Factor C to Me_2SO (FIG. **25**), which completely but reversibly inhibited the activation of Factor C, causing steric hindrance which renders it non-functional (Navas et al., 1990. *Biochem. Intl.* 21, 805–813). Similarly, EDTA chelates Ca^{2+}, which renders it unavailable to act as co-factor for the activation of Factor C (Dewanjess et al., 1990. *J. Nucl. Med.* 31, 234–245). The endotoxin receptors in both zymogen forms are probably similar to those typical of other LPS-binding proteins (Wright et al., 1986. *J. Exp. Med.* 164, 1876–188; Lei et al., 1988. *J. Immunol.* 141, 996–1011; Tobias et al., 1989. *J. Biol. Chem.* 264, 10867–10871). When either form of Factor C was complexed to LPS strips and incubated with its fluorometric substrate, the Factor C activity remained detectable. This shows that Factor C has an endotoxin-binding site which is unique from its serine protease catalytic site. This endotoxin receptor site is also capable of binding Me_2SO, which concomitantly results in its apparent inactivation.

Therefore, using a simple and unique protocol, two forms of Factor C were simultaneously isolated from the Carcinoscorpius amoebocyte lysate. This represents a considerable advancement over earlier methods (Nakamura et al., 1986. *Eur. J. Biochem.* 154, 511–521; Tokunaga et al., 1987. *Eur. J. Biochem.* 167, 405–416) used to isolate the two forms of Factor C from *T. tridentatus*. Furthermore, incorporation of 5% Me_2SO and 1 mM Na_2EDTA not only averted the endotoxin-induced activation of the enzymes, but was instrumental in the isolation, detection, and differentiation of one form of Factor C from the other.

Thus, during the initial stages of amoebocyte lysate preparation, Me_2SO can be used as a reversible inhibitor of endotoxin-induced Factor C activation. For this purpose, Me_2SO can be employed at a concentration of from about 5% to about 30%, v/v. At the final step of lysate preparation, this transient inactivation of Factor C can be reversed by removal of Me_2SO. This method will help to overcome the deleterious effects of the ubiquity of endotoxin, and ensures higher quality lysate preparations. Of course, Factor C preparations or solutions of any type can also be protected in this manner as well.

For example, Factor C can be maintained in its zymogen form in any type of preparation, including crude amoebocyte lysates, or by lysing amoebocytes in a solution comprising Me_2SO. This solution can further contain a chelating agent such as EDTA, where the lysing solution is pyrogen-free water, or any other solution suitable for lysis, suspension, or storage. Me_2SO can be employed in a concentration of from about 5% to about 30%, v/v, preferably about 5%, v/v, and when present, the chelating agent such as EDTA, Na_2EDTA, EGTA, or any other similarly acting chelating agent, can be employed at a concentration of about 1 mM. For long-term storage of such protected lysates before use, the lysates thus prepared can be lyophilized. These steps are useful for long-term storage and maintenance of industrial quantities of amoebocyte lysates prior to and during further treatment of the lysates, or during handling or storage of any Factor C-containing solutions or preparations prior to use to protect Factor C from autoactivation and to maintain it in its zymogen form. Me_2SO and EDTA can subsequently be removed by desalting through Sephadex G-25 or by any other chemical methods of neutralizing Me_2SO and EDTA when the enriched lysate is required, e.g., for LAL assay.

The invention being thus described, it will be obvious that the same may be varied in many ways. Such variations are not to be regarded as a departure from the spirit and scope of the invention, and all such modifications as would be obvious to one skilled in the art are intended to be included within the scope of the following claims.

Relevant Publications on *C. rotundicauda* Factor C

1. Ho, B. (1983) "An improved Limulus amoebocyte lysate assay". *Microbios Letts.* 25:81–84.
2. Ho, B. & Ding, J. L. (1985) "Comparison of sensitivity of Tachypleus and Limulus amoebocyte lysate in rapid detection of Gram negative bacteria". *Intl. Congr. on Microbiol. in the* 80's, eds. G. Lim & B. H. Nga, pp. 664–669.
3. Kim, J. C., Ding, J. L. & Ho, B. (1987) "Preparation of active amoebocyte lysate from Tachypleus gigas and *Carcinoscorpius rotundicauda*". Abstract in *Proc. Ann. Sci. Meeting of the Singapore Soc. Microbiol.* 10–11 Jan. 1987, ed. Y. C. Chan, p. 21.
4. Ding, J. L., Kim, J. C. & Ho, B. (1988) "Pokeweed mitogen stimulates DNA synthesis in cultured amoebocytes of *Carcinoscorpius rotundicauda*." Cytobios 55:147–154.
5. Kim, J. C., Ding, J. L. & Ho, B. (1988) "Endotoxin activation of clotting proteins from *Carcinoscorpius rotundicauda* amoebocyte lysate." In: *Advances in Biochem. & Biotechnol. in Asia and Oceania*, eds. A. Sipat, K. Ampon, R. Perumal, S. Aziz, and V. Thambyrajah, Malaysia, p. F1.
6. Navas III, M. A. A., Ding, J. L. & Ho, B. (1989) "Purification and characterisation of Factor C and proclotting enzyme from amoebocyte lysate of *Carcinoscorpius rotundicauda*". Abstract in *Proc. 5th FAOB Congr.*, 13–18 Aug. 1989, Seoul, S. Korea. p. 199.
7. Yeo, S. A., Ho, B. & Ding, J. L. (1989) "Preservation of Factor C activity and removal of its inhibitory factor from *Carcinoscorpius rotundicauda*." Abstract in *Intl. Symp. New Frontiers Food & Med. Microbiol.*, Singapore. p. 79.
8. Navas III, M. A. A., Ding, J. L. & Ho, B. (1990) "Inactivation of Factor C by dimethyl sulfoxide inhibits coagulation of the Carcinoscorpius amoebocyte lysate" *Biochem. Intl.*, Academic Press, Australia. 21(5) :805–813.
9. Ho, B., Kim. J. C. & Ding J. L. (1993) "Electrophoretic analysis of endotoxin-activated gelation reaction of *Carcinoscorpius rotundicauda* amoebocyte lysate" *Biochem. & Mol. Biol. Intl.*, Acad. Press, Australia. 29(4):687–694.
10. Ding, J. L., Navas, M. A. A. & Ho, B. (1993) "Two forms of Factor C from the amoebocytes of *Carcinoscorpius rotundicauda*: Purification and characterisation" *Biochim. et Biophys. Acta* 1202: 149–156.
11. Ding, J. L., Sababathy, Tr.K. & Ho, B. (1993) "Morphological changes in *Carcinoscorpius rotundicauda* amoebocytes and *E. coli* during their interaction" Cytobios 75:21–32.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 4

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4182 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: both

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Carcinoscorpius rotundicauda

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 569..3817

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GTATTTAATG TCTCAACGGT AAAGGTTTCA TTGTAGCTAA TATTTAACTT CCTCCCTGTG      60

CCCCAAATCG CGAGTATGAC GTCAGTTAAG ACTTCGTATT TTAAGAGTTA AACACGAGCC     120

TTAAAGAGCG ATATTTTTTT TGTTAAACAC TTCCAACTTA ATACAATTGG CAAACTTTCA     180

AAAATAAAGT GGAAAAGGAG GTAAAAAAGA TGAAAAAAAT TCGCATACAA TAGAATACAA     240

TAAAATGTGT TGTCTTTACT GTCAACACTT ACTGTTCGTT CGGTCACAGC TGTGAATCGG     300

GGTGACTTTA TGTTTGTAGT GGTCTTAAAA ACGGGTACTT GGTTGTTTTG AAAATTTTAA     360

AACCTACATA TGATTCTCCT AAAATTTTGT TTATAAATTA GCACCATTTG CGACCTAAAT     420

CTTTTTTGTA GTCTTAAGTT TAGTTGACAT AAAAACAAAA TTTGTAACAA CACACGGTAT     480

AAACTAAATA GCTTCAGATG GGTCGTATGA CAAGGAAACT TTTAAATAAT TATGAAAGTT     540

TTTTTAAAAT TTGACTAAGG TTTAGATT ATG TGG GTG ACA TGC TTC GAC ACG         592
                               Met Trp Val Thr Cys Phe Asp Thr
                                 1               5

TTT CTT TTT GTT TGT GAA AGT TCA GTT TTC TGT TTG TTG TGT GTG TGG        640
Phe Leu Phe Val Cys Glu Ser Ser Val Phe Cys Leu Leu Cys Val Trp
      10                  15                  20

AGG TTT GGT TTC TGT AGG TGG CGT GTT TTC TAC AGT TTT CCA TTC GTT        688
Arg Phe Gly Phe Cys Arg Trp Arg Val Phe Tyr Ser Phe Pro Phe Val
 25                  30                  35                  40

AAG TCA ACA GTT GTT TTA TTA CAG TGT TAC CAT TAC TCT CTC CAC AAT        736
Lys Ser Thr Val Val Leu Leu Gln Cys Tyr His Tyr Ser Leu His Asn
                 45                  50                  55

ACC TCA AAG TTC TAC TCT GTG AAT CCT GAC AAG CCA GAG TAC ATT CTT        784
Thr Ser Lys Phe Tyr Ser Val Asn Pro Asp Lys Pro Glu Tyr Ile Leu
           60                  65                  70

TCA GGT TTA GTT CTA GGG CTA CTA GCC CAA AAA ATG CGC CCA GTT CAG        832
Ser Gly Leu Val Leu Gly Leu Leu Ala Gln Lys Met Arg Pro Val Gln
      75                  80                  85

TCC AAA GGA GTA GAT CTA GGC TTG TGT GAT GAA ACG AGG TTC GAG TGT        880
Ser Lys Gly Val Asp Leu Gly Leu Cys Asp Glu Thr Arg Phe Glu Cys
 90                  95                  100

AAG TGT GGC GAT CCA GGC TAT GTG TTC AAC ATT CCA GTG AAA CAA TGT        928
Lys Cys Gly Asp Pro Gly Tyr Val Phe Asn Ile Pro Val Lys Gln Cys
105                 110                 115                 120
```

-continued

```
ACA TAC TTT TAT CGA TGG AGG CCG TAT TGT AAA CCA TGT GAT GAC CTG          976
Thr Tyr Phe Tyr Arg Trp Arg Pro Tyr Cys Lys Pro Cys Asp Asp Leu
                125                 130                 135

GAG GCT AAG GAT ATT TGT CCA AAG TAC AAA CGA TGT CAA GAG TGT AAG         1024
Glu Ala Lys Asp Ile Cys Pro Lys Tyr Lys Arg Cys Gln Glu Cys Lys
            140                 145                 150

GCT GGT CTT GAT AGT TGT GTT ACT TGT CCA CCT AAC AAA TAT GGT ACT         1072
Ala Gly Leu Asp Ser Cys Val Thr Cys Pro Pro Asn Lys Tyr Gly Thr
        155                 160                 165

TGG TGT AGC GGT GAA TGT CAG TGT AAG AAT GGA GGT ATC TGT GAC CAG         1120
Trp Cys Ser Gly Glu Cys Gln Cys Lys Asn Gly Gly Ile Cys Asp Gln
    170                 175                 180

AGG ACA GGA GCT TGT GCA TGT CGT GAC AGA TAT GAA GGG GTG CAC TGT         1168
Arg Thr Gly Ala Cys Ala Cys Arg Asp Arg Tyr Glu Gly Val His Cys
185                 190                 195                 200

GAA ATT CTC AAA GGT TGT CCT CTT CTT CCA TCG GAT TCT CAG GTT CAG         1216
Glu Ile Leu Lys Gly Cys Pro Leu Leu Pro Ser Asp Ser Gln Val Gln
                205                 210                 215

GAA GTC AGA AAT CCA CCA GAT AAT CCC CAA ACT ATT GAC TAC AGC TGT         1264
Glu Val Arg Asn Pro Pro Asp Asn Pro Gln Thr Ile Asp Tyr Ser Cys
            220                 225                 230

TCA CCA GGG TTC AAG CTT AAG GGT ATG GCA CGA ATT AGC TGT CTC CCA         1312
Ser Pro Gly Phe Lys Leu Lys Gly Met Ala Arg Ile Ser Cys Leu Pro
        235                 240                 245

AAT GGA CAG TGG AGT AAC TTT CCA CCC AAA TGT ATT CGA GAA TGT GCC         1360
Asn Gly Gln Trp Ser Asn Phe Pro Pro Lys Cys Ile Arg Glu Cys Ala
    250                 255                 260

ATG GTT TCA TCT CCA GAA CAT GGG AAA GTG AAT GCT CTT AGT GGT GAT         1408
Met Val Ser Ser Pro Glu His Gly Lys Val Asn Ala Leu Ser Gly Asp
265                 270                 275                 280

ATG ATA GAA GGG GCT ACT TTA CGG TTC TCA TGT GAT AGT CCC TAC TAC         1456
Met Ile Glu Gly Ala Thr Leu Arg Phe Ser Cys Asp Ser Pro Tyr Tyr
                285                 290                 295

TTG ATT GGT CAA GAA ACA TTA ACC TGT CAG GGT AAT GGT CAG TGG AAT         1504
Leu Ile Gly Gln Glu Thr Leu Thr Cys Gln Gly Asn Gly Gln Trp Asn
            300                 305                 310

GGA CAG ATA CCA CAA TGT AAG AAC TTA GTC TTC TGT CCT GAC CTG GAT         1552
Gly Gln Ile Pro Gln Cys Lys Asn Leu Val Phe Cys Pro Asp Leu Asp
        315                 320                 325

CCT GTA AAC CAT GCT GAA CAC AAG GTT AAA ATT GGT GTG GAA CAA AAA         1600
Pro Val Asn His Ala Glu His Lys Val Lys Ile Gly Val Glu Gln Lys
    330                 335                 340

TAT GGT CAG TTT CCT CAA GGC ACT GAA GTG ACC TAT ACG TGT TCG GGT         1648
Tyr Gly Gln Phe Pro Gln Gly Thr Glu Val Thr Tyr Thr Cys Ser Gly
345                 350                 355                 360

AAC TAC TTC TTG ATG GGT TTT GAC ACC TTA AAA TGT AAC CCT GAT GGG         1696
Asn Tyr Phe Leu Met Gly Phe Asp Thr Leu Lys Cys Asn Pro Asp Gly
                365                 370                 375

TCT TGG TCA GGA TCA CAG CCA TCC TGT GTT AAA GTG GCA GAC AGA GAG         1744
Ser Trp Ser Gly Ser Gln Pro Ser Cys Val Lys Val Ala Asp Arg Glu
            380                 385                 390

GTC GAC TGT GAC AGT AAA GCT GTA GAC TTC TTG GAT GAT GTT GGT GAA         1792
Val Asp Cys Asp Ser Lys Ala Val Asp Phe Leu Asp Asp Val Gly Glu
        395                 400                 405

CCT GTC AGG ATC CAC TGT CCT GCT GGC TGT TCT TTG ACA GCT GGT ACT         1840
Pro Val Arg Ile His Cys Pro Ala Gly Cys Ser Leu Thr Ala Gly Thr
    410                 415                 420

GTG TGG GGT ACA GCC ATA TAC CAT GAA CTT TCC TCA GTG TGT CGT GCA         1888
Val Trp Gly Thr Ala Ile Tyr His Glu Leu Ser Ser Val Cys Arg Ala
425                 430                 435                 440
```

-continued

```
GCC ATC CAT GCT GGC AAG CTT CCA AAC TCT GGA GGA GCG GTG CAT GTT          1936
Ala Ile His Ala Gly Lys Leu Pro Asn Ser Gly Gly Ala Val His Val
                445                 450                 455

GTG AAC AAT GGC CCC TAC TCG GAC TTT CTG GGT AGT GAC CTG AAT GGG          1984
Val Asn Asn Gly Pro Tyr Ser Asp Phe Leu Gly Ser Asp Leu Asn Gly
            460                 465                 470

ATA AAA TCC GAA GAG TTG AAG TCT CTT GCC CGG AGT TTC CGA TTC GAT          2032
Ile Lys Ser Glu Glu Leu Lys Ser Leu Ala Arg Ser Phe Arg Phe Asp
        475                 480                 485

TAT GTC AGT TCC TCC TCA GCA GGT AAA TCA GGA TGT CCT GAT GGA TGG          2080
Tyr Val Ser Ser Ser Thr Ala Gly Lys Ser Gly Cys Pro Asp Gly Trp
    490                 495                 500

TTT GAG GTA GAC GAG AAC TGT GTG TAC GTT ACA TCA AAA CAG AGA GCC          2128
Phe Glu Val Asp Glu Asn Cys Val Tyr Val Thr Ser Lys Gln Arg Ala
505                 510                 515                 520

TGG GAA AGA GCT CAA GGT GTG TGT ACC AAT ATG GCT GCT CGT CTT GCT          2176
Trp Glu Arg Ala Gln Gly Val Cys Thr Asn Met Ala Ala Arg Leu Ala
                525                 530                 535

GTG CTG GAC AAA GAT GTA ATT CCA AAT TCA TTG ACT GAG ACT CTA CGA          2224
Val Leu Asp Lys Asp Val Ile Pro Asn Ser Leu Thr Glu Thr Leu Arg
            540                 545                 550

GGG AAA GGG TTA ACA ACC ACG TGG ATA GGA TTG CAC AGA CTA GAT GCT          2272
Gly Lys Gly Leu Thr Thr Thr Trp Ile Gly Leu His Arg Leu Asp Ala
        555                 560                 565

GAG AAG CCC TTT ATT TGG GAG TTA ATG GAT CGT AGT AAT GTG GTT CTG          2320
Glu Lys Pro Phe Ile Trp Glu Leu Met Asp Arg Ser Asn Val Val Leu
    570                 575                 580

AAT GAT AAC CTA ACA TTC TGG GCC TCT GGC GAA CCT GGA AAT GAA ACT          2368
Asn Asp Asn Leu Thr Phe Trp Ala Ser Gly Glu Pro Gly Asn Glu Thr
585                 590                 595                 600

AAC TGT GTA TAT ATG GAC ATC CAA GAT CAG TTG CAG TCT GTG TGG AAA          2416
Asn Cys Val Tyr Met Asp Ile Gln Asp Gln Leu Gln Ser Val Trp Lys
                605                 610                 615

ACC AAG TCA TGT TTT CAG CCC TCA AGT TTT GCT TGC ATG ATG GAT CTG          2464
Thr Lys Ser Cys Phe Gln Pro Ser Ser Phe Ala Cys Met Met Asp Leu
            620                 625                 630

TCA GAC AGA AAT AAA GCC AAA TGC GAT GAT CCT GGA TCA CTG GAA AAT          2512
Ser Asp Arg Asn Lys Ala Lys Cys Asp Asp Pro Gly Ser Leu Glu Asn
        635                 640                 645

GGA CAC GCC ACA CTT CAT GGA CAA AGT ATT GAT GGG TTC TAT GCT GGT          2560
Gly His Ala Thr Leu His Gly Gln Ser Ile Asp Gly Phe Tyr Ala Gly
    650                 655                 660

TCT TCT ATA AGG TAC AGC TGT GAG GTT CTC CAC TAC CTC AGT GGA ACT          2608
Ser Ser Ile Arg Tyr Ser Cys Glu Val Leu His Tyr Leu Ser Gly Thr
665                 670                 675                 680

GAA ACC GTA ACT TGT ACA ACA AAT GGC ACA TGG AGT GCT CCT AAA CCT          2656
Glu Thr Val Thr Cys Thr Thr Asn Gly Thr Trp Ser Ala Pro Lys Pro
                685                 690                 695

CGA TGT ATC AAA GTC ATC ACC TGC CAA AAC CCC CCT GTA CCA TCA TAT          2704
Arg Cys Ile Lys Val Ile Thr Cys Gln Asn Pro Pro Val Pro Ser Tyr
            700                 705                 710

GGT TCT GTG GAA ATC AAA CCC CCA AGT CGG ACA AAC TCG ATA AGT CGT          2752
Gly Ser Val Glu Ile Lys Pro Pro Ser Arg Thr Asn Ser Ile Ser Arg
        715                 720                 725

GTT GGG TCA CCT TTC TTG AGG TTG CCA CGG TTA CCC CTC CCA TTA GCC          2800
Val Gly Ser Pro Phe Leu Arg Leu Pro Arg Leu Pro Leu Pro Leu Ala
    730                 735                 740

AGA GCA GCC AAA CCT CCT CCA AAA CCT AGA TCC TCA CAA CCC TCT ACT          2848
Arg Ala Ala Lys Pro Pro Pro Lys Pro Arg Ser Ser Gln Pro Ser Thr
745                 750                 755                 760
```

-continued

```
GTG GAC TTG GCT TCT AAA GTT AAA CTA CCT GAA GGT CAT TAC CGG GTA        2896
Val Asp Leu Ala Ser Lys Val Lys Leu Pro Glu Gly His Tyr Arg Val
                765                 770                 775

GGG TCT CGA GCC ATT TAC ACG TGC GAG TCG AGA TAC TAC GAA CTA CTT        2944
Gly Ser Arg Ala Ile Tyr Thr Cys Glu Ser Arg Tyr Tyr Glu Leu Leu
            780                 785                 790

GGA TCT CAA GGC AGA AGA TGT GAC TCT AAT GGA AAC TGG AGT GGT CGG        2992
Gly Ser Gln Gly Arg Arg Cys Asp Ser Asn Gly Asn Trp Ser Gly Arg
        795                 800                 805

CCA GCG AGC TGT ATT CCA GTT TGT GGA CGG TCA GAC TCT CCT CGT TCT        3040
Pro Ala Ser Cys Ile Pro Val Cys Gly Arg Ser Asp Ser Pro Arg Ser
    810                 815                 820

CCT TTT ATC TGG AAT GGG AAT TCT ACA GAA ATA GGT CAG TGG CCG TGG        3088
Pro Phe Ile Trp Asn Gly Asn Ser Thr Glu Ile Gly Gln Trp Pro Trp
825                 830                 835                 840

CAG GCA GGA ATC TCT AGA TGG CTT GCA GAC CAC AAT ATG TGG TTT CTC        3136
Gln Ala Gly Ile Ser Arg Trp Leu Ala Asp His Asn Met Trp Phe Leu
                845                 850                 855

CAG TGT GGA GGA TCT CTA TTG AAT GAG AAA TGG ATC GTC ACT GCT GCC        3184
Gln Cys Gly Gly Ser Leu Leu Asn Glu Lys Trp Ile Val Thr Ala Ala
            860                 865                 870

CAC TGT GTC ACC TAC TCT GCT ACT GCT GAG ATT ATT GAC CCC AAT CAG        3232
His Cys Val Thr Tyr Ser Ala Thr Ala Glu Ile Ile Asp Pro Asn Gln
        875                 880                 885

TTT AAA ATG TAT CTG GGC AAG TAC TAC CGT GAT GAC AGT AGA GAC GAT        3280
Phe Lys Met Tyr Leu Gly Lys Tyr Tyr Arg Asp Asp Ser Arg Asp Asp
    890                 895                 900

GAC TAT GTA CAA GTA AGA GAG GCT CTT GAG ATC CAC GTG AAT CCT AAC        3328
Asp Tyr Val Gln Val Arg Glu Ala Leu Glu Ile His Val Asn Pro Asn
905                 910                 915                 920

TAC GAC CCC GGC AAT CTC AAC TTT GAC ATA GCC CTA ATT CAA CTG AAA        3376
Tyr Asp Pro Gly Asn Leu Asn Phe Asp Ile Ala Leu Ile Gln Leu Lys
                925                 930                 935

ACT CCT GTT ACT TTG ACA ACA CGA GTC CAA CCA ATC TGT CTG CCT ACT        3424
Thr Pro Val Thr Leu Thr Thr Arg Val Gln Pro Ile Cys Leu Pro Thr
            940                 945                 950

GAC ATC ACA ACA AGA GAA CAC TTG AAG GAG GGA ACA TTA GCA GTG GTG        3472
Asp Ile Thr Thr Arg Glu His Leu Lys Glu Gly Thr Leu Ala Val Val
        955                 960                 965

ACA GGT TGG GGT TTG AAT GAA AAC AAC ACC TAT TCA GAG ACG ATT CAA        3520
Thr Gly Trp Gly Leu Asn Glu Asn Asn Thr Tyr Ser Glu Thr Ile Gln
    970                 975                 980

CAA GCT GTG CTA CCT GTT GTT GCA GCC AGC ACC TGT GAA GAG GGG TAC        3568
Gln Ala Val Leu Pro Val Val Ala Ala Ser Thr Cys Glu Glu Gly Tyr
985                 990                 995                 1000

AAG GAA GCA GAC TTA CCA CTG ACA GTA ACA GAG AAC ATG TTC TGT GCA        3616
Lys Glu Ala Asp Leu Pro Leu Thr Val Thr Glu Asn Met Phe Cys Ala
                1005                1010                1015

GGT TAC AAG AAG GGA CGT TAT GAT GCC TGC AGT GGG GAC AGT GGA GGA        3664
Gly Tyr Lys Lys Gly Arg Tyr Asp Ala Cys Ser Gly Asp Ser Gly Gly
            1020                1025                1030

CCT TTA GTG TTT GCT GAT GAT TCC CGT ACC GAA AGG CGG TGG GTC TTG        3712
Pro Leu Val Phe Ala Asp Asp Ser Arg Thr Glu Arg Arg Trp Val Leu
        1035                1040                1045

GAA GGG ATT GTC AGC TGG GGC AGT CCC AGT GGA TGT GGC AAG GCG AAC        3760
Glu Gly Ile Val Ser Trp Gly Ser Pro Ser Gly Cys Gly Lys Ala Asn
    1050                1055                1060

CAG TAC GGG GGC TTC ACT AAA GTT AAC GTT TTC CTG TCA TGG ATT AGG        3808
Gln Tyr Gly Gly Phe Thr Lys Val Asn Val Phe Leu Ser Trp Ile Arg
1065                1070                1075                1080
```

-continued

```
CAG TTC ATT TGAAACTGAT CTAAATATTT TAAGCATGGT TATAAACGTC                3857
Gln Phe Ile

TTGTTCCTAT TATTGCTTTA CTGGTTTAAC CCATAAGAAG GTTAACGGGG TAAGGCACAA     3917

GGATCATTGT TTCTGTTTGT TTTTACAAAT GGTTCTTTTA GTCAGTGAAT GAGAATAGTA     3977

TCCATTGGAG ACTGTTACCT TTTATTCTAC CTTTTTATAT TACTATGCAA GTATTTGGGA     4037

TATCTTCTAC ACATGAAAAT TCTGTCATTT TACCATAAAT TTGGTTTCTG GTGTGTGTGT     4097

TAAGTCCACC ACTAGAGAAC GATGTAATTT TCAATAGTAC ATGAAATAAA TATAGAACAA     4157

ATCTATTATA AAAAAAAAAA AAAAA                                           4182
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1083 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Trp Val Thr Cys Phe Asp Thr Phe Leu Phe Val Cys Glu Ser Ser
 1               5                  10                  15

Val Phe Cys Leu Leu Cys Val Trp Arg Phe Gly Phe Cys Arg Trp Arg
            20                  25                  30

Val Phe Tyr Ser Phe Pro Phe Val Lys Ser Thr Val Val Leu Leu Gln
        35                  40                  45

Cys Tyr His Tyr Ser Leu His Asn Thr Ser Lys Phe Tyr Ser Val Asn
    50                  55                  60

Pro Asp Lys Pro Glu Tyr Ile Leu Ser Gly Leu Val Leu Gly Leu Leu
65                  70                  75                  80

Ala Gln Lys Met Arg Pro Val Gln Ser Lys Gly Val Asp Leu Gly Leu
                85                  90                  95

Cys Asp Glu Thr Arg Phe Glu Cys Lys Cys Gly Asp Pro Gly Tyr Val
            100                 105                 110

Phe Asn Ile Pro Val Lys Gln Cys Thr Tyr Phe Tyr Arg Trp Arg Pro
        115                 120                 125

Tyr Cys Lys Pro Cys Asp Asp Leu Glu Ala Lys Asp Ile Cys Pro Lys
    130                 135                 140

Tyr Lys Arg Cys Gln Glu Cys Lys Ala Gly Leu Asp Ser Cys Val Thr
145                 150                 155                 160

Cys Pro Pro Asn Lys Tyr Gly Thr Trp Cys Ser Gly Glu Cys Gln Cys
                165                 170                 175

Lys Asn Gly Gly Ile Cys Asp Gln Arg Thr Gly Ala Cys Ala Cys Arg
            180                 185                 190

Asp Arg Tyr Glu Gly Val His Cys Glu Ile Leu Lys Gly Cys Pro Leu
        195                 200                 205

Leu Pro Ser Asp Ser Gln Val Gln Glu Val Arg Asn Pro Pro Asp Asn
    210                 215                 220

Pro Gln Thr Ile Asp Tyr Ser Cys Ser Pro Gly Phe Lys Leu Lys Gly
225                 230                 235                 240

Met Ala Arg Ile Ser Cys Leu Pro Asn Gly Gln Trp Ser Asn Phe Pro
                245                 250                 255

Pro Lys Cys Ile Arg Glu Cys Ala Met Val Ser Ser Pro Glu His Gly
            260                 265                 270

Lys Val Asn Ala Leu Ser Gly Asp Met Ile Glu Gly Ala Thr Leu Arg
```

Seikagaku Corporation Exhibit 2004 - Page 284 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

-continued

```
                275                 280                 285

Phe Ser Cys Asp Ser Pro Tyr Tyr Leu Ile Gly Gln Glu Thr Leu Thr
    290                 295                 300

Cys Gln Gly Asn Gly Gln Trp Asn Gly Gln Ile Pro Gln Cys Lys Asn
305                 310                 315                 320

Leu Val Phe Cys Pro Asp Leu Asp Pro Val Asn His Ala Glu His Lys
                325                 330                 335

Val Lys Ile Gly Val Glu Gln Lys Tyr Gly Gln Phe Pro Gln Gly Thr
            340                 345                 350

Glu Val Thr Tyr Thr Cys Ser Gly Asn Tyr Phe Leu Met Gly Phe Asp
        355                 360                 365

Thr Leu Lys Cys Asn Pro Asp Gly Ser Trp Ser Gly Ser Gln Pro Ser
    370                 375                 380

Cys Val Lys Val Ala Asp Arg Glu Val Asp Cys Asp Ser Lys Ala Val
385                 390                 395                 400

Asp Phe Leu Asp Asp Val Gly Glu Pro Val Arg Ile His Cys Pro Ala
                405                 410                 415

Gly Cys Ser Leu Thr Ala Gly Thr Val Trp Gly Thr Ala Ile Tyr His
            420                 425                 430

Glu Leu Ser Ser Val Cys Arg Ala Ala Ile His Ala Gly Lys Leu Pro
        435                 440                 445

Asn Ser Gly Gly Ala Val His Val Val Asn Asn Gly Pro Tyr Ser Asp
    450                 455                 460

Phe Leu Gly Ser Asp Leu Asn Gly Ile Lys Ser Glu Glu Leu Lys Ser
465                 470                 475                 480

Leu Ala Arg Ser Phe Arg Phe Asp Tyr Val Ser Ser Ser Thr Ala Gly
                485                 490                 495

Lys Ser Gly Cys Pro Asp Gly Trp Phe Glu Val Asp Glu Asn Cys Val
            500                 505                 510

Tyr Val Thr Ser Lys Gln Arg Ala Trp Glu Arg Ala Gln Gly Val Cys
        515                 520                 525

Thr Asn Met Ala Ala Arg Leu Ala Val Leu Asp Lys Asp Val Ile Pro
    530                 535                 540

Asn Ser Leu Thr Glu Thr Leu Arg Gly Lys Gly Leu Thr Thr Thr Trp
545                 550                 555                 560

Ile Gly Leu His Arg Leu Asp Ala Glu Lys Pro Phe Ile Trp Glu Leu
                565                 570                 575

Met Asp Arg Ser Asn Val Val Leu Asn Asp Asn Leu Thr Phe Trp Ala
            580                 585                 590

Ser Gly Glu Pro Gly Asn Glu Thr Asn Cys Val Tyr Met Asp Ile Gln
        595                 600                 605

Asp Gln Leu Gln Ser Val Trp Lys Thr Lys Ser Cys Phe Gln Pro Ser
    610                 615                 620

Ser Phe Ala Cys Met Met Asp Leu Ser Asp Arg Asn Lys Ala Lys Cys
625                 630                 635                 640

Asp Asp Pro Gly Ser Leu Glu Asn Gly His Ala Thr Leu His Gly Gln
                645                 650                 655

Ser Ile Asp Gly Phe Tyr Ala Gly Ser Ser Ile Arg Tyr Ser Cys Glu
            660                 665                 670

Val Leu His Tyr Leu Ser Gly Thr Glu Thr Val Thr Cys Thr Thr Asn
        675                 680                 685

Gly Thr Trp Ser Ala Pro Lys Pro Arg Cys Ile Lys Val Ile Thr Cys
    690                 695                 700
```

-continued

```
Gln Asn Pro Pro Val Pro Ser Tyr Gly Ser Val Glu Ile Lys Pro Pro
705                 710                 715                 720

Ser Arg Thr Asn Ser Ile Ser Arg Val Gly Ser Pro Phe Leu Arg Leu
                725                 730                 735

Pro Arg Leu Pro Leu Pro Leu Ala Arg Ala Ala Lys Pro Pro Pro Lys
            740                 745                 750

Pro Arg Ser Ser Gln Pro Ser Thr Val Asp Leu Ala Ser Lys Val Lys
        755                 760                 765

Leu Pro Glu Gly His Tyr Arg Val Gly Ser Arg Ala Ile Tyr Thr Cys
    770                 775                 780

Glu Ser Arg Tyr Tyr Glu Leu Leu Gly Ser Gln Gly Arg Arg Cys Asp
785                 790                 795                 800

Ser Asn Gly Asn Trp Ser Gly Arg Pro Ala Ser Cys Ile Pro Val Cys
                805                 810                 815

Gly Arg Ser Asp Ser Pro Arg Ser Pro Phe Ile Trp Asn Gly Asn Ser
            820                 825                 830

Thr Glu Ile Gly Gln Trp Pro Trp Gln Ala Gly Ile Ser Arg Trp Leu
        835                 840                 845

Ala Asp His Asn Met Trp Phe Leu Gln Cys Gly Gly Ser Leu Leu Asn
    850                 855                 860

Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Thr Tyr Ser Ala Thr
865                 870                 875                 880

Ala Glu Ile Ile Asp Pro Asn Gln Phe Lys Met Tyr Leu Gly Lys Tyr
                885                 890                 895

Tyr Arg Asp Asp Ser Arg Asp Asp Asp Tyr Val Gln Val Arg Glu Ala
            900                 905                 910

Leu Glu Ile His Val Asn Pro Asn Tyr Asp Pro Gly Asn Leu Asn Phe
        915                 920                 925

Asp Ile Ala Leu Ile Gln Leu Lys Thr Pro Val Thr Leu Thr Thr Arg
    930                 935                 940

Val Gln Pro Ile Cys Leu Pro Thr Asp Ile Thr Thr Arg Glu His Leu
945                 950                 955                 960

Lys Glu Gly Thr Leu Ala Val Val Thr Gly Trp Gly Leu Asn Glu Asn
                965                 970                 975

Asn Thr Tyr Ser Glu Thr Ile Gln Gln Ala Val Leu Pro Val Val Ala
            980                 985                 990

Ala Ser Thr Cys Glu Glu Gly Tyr Lys Glu Ala Asp Leu Pro Leu Thr
        995                 1000                1005

Val Thr Glu Asn Met Phe Cys Ala Gly Tyr Lys Lys Gly Arg Tyr Asp
    1010                1015                1020

Ala Cys Ser Gly Asp Ser Gly Gly Pro Leu Val Phe Ala Asp Asp Ser
1025                1030                1035                1040

Arg Thr Glu Arg Arg Trp Val Leu Glu Gly Ile Val Ser Trp Gly Ser
                1045                1050                1055

Pro Ser Gly Cys Gly Lys Ala Asn Gln Tyr Gly Gly Phe Thr Lys Val
            1060                1065                1070

Asn Val Phe Leu Ser Trp Ile Arg Gln Phe Ile
        1075                1080
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: both

Seikagaku Corporation Exhibit 2004 - Page 286 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

-continued

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Carcinoscorpius rotundicauda

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 18..3074

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GTGAAGGTAA CTTAAGT ATG GTC TTA GCG TCG TTT TTG GTG TCT GGT TTA          50
                    Met Val Leu Ala Ser Phe Leu Val Ser Gly Leu
                      1               5                  10

GTT CTA GGG CTA CTA GCC CAA AAA ATG CGC CCA GTT CAG TCC AAA GGA          98
Val Leu Gly Leu Leu Ala Gln Lys Met Arg Pro Val Gln Ser Lys Gly
            15                  20                  25

GTA GAT CTA GGC TTG TGT GAT GAA ACG AGG TTC GAG TGT AAG TGT GGC         146
Val Asp Leu Gly Leu Cys Asp Glu Thr Arg Phe Glu Cys Lys Cys Gly
        30                  35                  40

GAT CCA GGC TAT GTG TTC AAC ATT CCA GTG AAA CAA TGT ACA TAC TTT         194
Asp Pro Gly Tyr Val Phe Asn Ile Pro Val Lys Gln Cys Thr Tyr Phe
    45                  50                  55

TAT CGA TGG AGG CCG TAT TGT AAA CCA TGT GAT GAC CTG GAG GCT AAG         242
Tyr Arg Trp Arg Pro Tyr Cys Lys Pro Cys Asp Asp Leu Glu Ala Lys
60                  65                  70                  75

GAT ATT TGT CCA AAG TAC AAA CGA TGT CAA GAG TGT AAG GCT GGT CTT         290
Asp Ile Cys Pro Lys Tyr Lys Arg Cys Gln Glu Cys Lys Ala Gly Leu
                80                  85                  90

GAT AGT TGT GTT ACT TGT CCA CCT AAC AAA TAT GGT ACT TGG TGT AGC         338
Asp Ser Cys Val Thr Cys Pro Pro Asn Lys Tyr Gly Thr Trp Cys Ser
            95                  100                 105

GGT GAA TGT CAG TGT AAG AAT GGA GGT ATC TGT GAC CAG AGG ACA GGA         386
Gly Glu Cys Gln Cys Lys Asn Gly Gly Ile Cys Asp Gln Arg Thr Gly
        110                 115                 120

GCT TGT GCA TGT CGT GAC AGA TAT GAA GGG GTG CAC TGT GAA ATT CTC         434
Ala Cys Ala Cys Arg Asp Arg Tyr Glu Gly Val His Cys Glu Ile Leu
    125                 130                 135

AAA GGT TGT CCT CTT CTT CCA TCG GAT TCT CAG GTT CAG GAA GTC AGA         482
Lys Gly Cys Pro Leu Leu Pro Ser Asp Ser Gln Val Gln Glu Val Arg
140                 145                 150                 155

AAT CCA CCA GAT AAT CCC CAA ACT ATT GAC TAC AGC TGT TCA CCA GGG         530
Asn Pro Pro Asp Asn Pro Gln Thr Ile Asp Tyr Ser Cys Ser Pro Gly
                160                 165                 170

TTC AAG CTT AAG GGT ATG GCA CGA ATT AGC TGT CTC CCA AAT GGA CAG         578
Phe Lys Leu Lys Gly Met Ala Arg Ile Ser Cys Leu Pro Asn Gly Gln
            175                 180                 185

TGG AGT AAC TTT CCA CCC AAA TGT ATT CGA GAA TGT GCC ATG GTT TCA         626
Trp Ser Asn Phe Pro Pro Lys Cys Ile Arg Glu Cys Ala Met Val Ser
        190                 195                 200

TCT CCA GAA CAT GGG AAA GTG AAT GCT CTT AGT GGT GAT ATG ATA GAA         674
Ser Pro Glu His Gly Lys Val Asn Ala Leu Ser Gly Asp Met Ile Glu
    205                 210                 215

GGG GCT ACT TTA CGG TTC TCA TGT GAT AGT CCC TAC TAC TTG ATT GGT         722
Gly Ala Thr Leu Arg Phe Ser Cys Asp Ser Pro Tyr Tyr Leu Ile Gly
220                 225                 230                 235

CAA GAA ACA TTA ACC TGT CAG GGT AAT GGT CAG TGG AAT GGA CAG ATA         770
Gln Glu Thr Leu Thr Cys Gln Gly Asn Gly Gln Trp Asn Gly Gln Ile
                240                 245                 250
```

-continued

```
CCA CAA TGT AAG AAC TTG GTC TTC TGT CCT GAC CTG GAT CCT GTA AAC                818
Pro Gln Cys Lys Asn Leu Val Phe Cys Pro Asp Leu Asp Pro Val Asn
            255                 260                 265

CAT GCT GAA CAC AAG GTT AAA ATT GGT GTG GAA CAA AAA TAT GGT CAG               866
His Ala Glu His Lys Val Lys Ile Gly Val Glu Gln Lys Tyr Gly Gln
        270                 275                 280

TTT CCT CAA GGC ACT GAA GTG ACC TAT ACG TGT TCG GGT AAC TAC TTC               914
Phe Pro Gln Gly Thr Glu Val Thr Tyr Thr Cys Ser Gly Asn Tyr Phe
    285                 290                 295

TTG ATG GGT TTT GAC ACC TTA AAA TGT AAC CCT GAT GGG TCT TGG TCA               962
Leu Met Gly Phe Asp Thr Leu Lys Cys Asn Pro Asp Gly Ser Trp Ser
300                 305                 310                 315

GGA TCA CAG CCA TCC TGT GTT AAA GTG GCA GAC AGA GAG GTC GAC TGT              1010
Gly Ser Gln Pro Ser Cys Val Lys Val Ala Asp Arg Glu Val Asp Cys
                320                 325                 330

GAC AGT AAA GCT GTA GAC TTC TTG GAT GAT GTT GGT GAA CCT GTC AGG              1058
Asp Ser Lys Ala Val Asp Phe Leu Asp Asp Val Gly Glu Pro Val Arg
            335                 340                 345

ATC CAC TGT CCT GCT GGC TGT TCT TTG ACA GCT GGT ACT GTG TGG GGT              1106
Ile His Cys Pro Ala Gly Cys Ser Leu Thr Ala Gly Thr Val Trp Gly
        350                 355                 360

ACA GCC ATA TAC CAT GAA CTT TCC TCA GTG TGT CGT GCA GCC ATC CAT              1154
Thr Ala Ile Tyr His Glu Leu Ser Ser Val Cys Arg Ala Ala Ile His
    365                 370                 375

GCT GGC AAG CTT CCA AAC TCT GGA GGA GCG GTG CAT GTT GTG AAC AAT              1202
Ala Gly Lys Leu Pro Asn Ser Gly Gly Ala Val His Val Val Asn Asn
380                 385                 390                 395

GGC CCC TAC TCG GAC TTT CTG GGT AGT GAC CTG AAT GGG ATA AAA TCG              1250
Gly Pro Tyr Ser Asp Phe Leu Gly Ser Asp Leu Asn Gly Ile Lys Ser
                400                 405                 410

GAA GAG TTG AAG TCT CTT GCC CGG AGT TTC CGA TTC GAT TAT GTC CGT              1298
Glu Glu Leu Lys Ser Leu Ala Arg Ser Phe Arg Phe Asp Tyr Val Arg
            415                 420                 425

TCC TCC ACA GCA GGT AAA TCA GGA TGT CCT GAT GGA TGG TTT GAG GTA              1346
Ser Ser Thr Ala Gly Lys Ser Gly Cys Pro Asp Gly Trp Phe Glu Val
        430                 435                 440

GAC GAG AAC TGT GTG TAC GTT ACA TCA AAA CAG AGA GCC TGG GAA AGA              1394
Asp Glu Asn Cys Val Tyr Val Thr Ser Lys Gln Arg Ala Trp Glu Arg
    445                 450                 455

GCT CAA GGT GTG TGT ACC AAT ATG GCT GCT CGT CTT GCT GTG CTG GAC              1442
Ala Gln Gly Val Cys Thr Asn Met Ala Ala Arg Leu Ala Val Leu Asp
460                 465                 470                 475

AAA GAT GTA ATT CCA AAT TCG TTG ACT GAG ACT CTA CGA GGG AAA GGG              1490
Lys Asp Val Ile Pro Asn Ser Leu Thr Glu Thr Leu Arg Gly Lys Gly
                480                 485                 490

TTA ACA ACC ACG TGG ATA GGA TTG CAC AGA CTA GAT GCT GAG AAG CCC              1538
Leu Thr Thr Thr Trp Ile Gly Leu His Arg Leu Asp Ala Glu Lys Pro
            495                 500                 505

TTT ATT TGG GAG TTA ATG GAT CGT AGT AAT GTG GTT CTG AAT GAT AAC              1586
Phe Ile Trp Glu Leu Met Asp Arg Ser Asn Val Val Leu Asn Asp Asn
        510                 515                 520

CTA ACA TTC TGG GCC TCT GGC GAA CCT GGA AAT GAA ACT AAC TGT GTA              1634
Leu Thr Phe Trp Ala Ser Gly Glu Pro Gly Asn Glu Thr Asn Cys Val
    525                 530                 535

TAT ATG GAC ATC CAA GAT CAG TTG CAG TCT GTG TGG AAA ACC AAG TCA              1682
Tyr Met Asp Ile Gln Asp Gln Leu Gln Ser Val Trp Lys Thr Lys Ser
540                 545                 550                 555

TGT TTT CAG CCC TCA AGT TTT GCT TGC ATG ATG GAT CTG TCA GAC AGA              1730
Cys Phe Gln Pro Ser Ser Phe Ala Cys Met Met Asp Leu Ser Asp Arg
                560                 565                 570
```

-continued

```
AAT AAA GCC AAA TGC GAT GAT CCT GGA TCA CTG GAA AAT GGA CAC GCC          1778
Asn Lys Ala Lys Cys Asp Asp Pro Gly Ser Leu Glu Asn Gly His Ala
                575                 580                 585

ACA CTT CAT GGA CAA AGT ATT GAT GGG TTC TAT GCT GGT TCT TCT ATA          1826
Thr Leu His Gly Gln Ser Ile Asp Gly Phe Tyr Ala Gly Ser Ser Ile
            590                 595                 600

AGG TAC AGC TGT GAG GTT CTC CAC TAC CTC AGT GGA ACT GAA ACC GTA          1874
Arg Tyr Ser Cys Glu Val Leu His Tyr Leu Ser Gly Thr Glu Thr Val
        605                 610                 615

ACT TGT ACA ACA AAT GGC ACA TGG AGT GCT CCT AAA CCT CGA TGT ATC          1922
Thr Cys Thr Thr Asn Gly Thr Trp Ser Ala Pro Lys Pro Arg Cys Ile
620                 625                 630                 635

AAA GTC ATC ACC TGC CAA AAC CCC CCT GTA CCA TCA TAT GGT TCT GTG          1970
Lys Val Ile Thr Cys Gln Asn Pro Pro Val Pro Ser Tyr Gly Ser Val
                640                 645                 650

GAA ATC AAA CCC CCA AGT CGG ACA AAC TCG ATA AGT CGT GTT GGG TCA          2018
Glu Ile Lys Pro Pro Ser Arg Thr Asn Ser Ile Ser Arg Val Gly Ser
            655                 660                 665

CCT TTC TTG AGG TTG CCA CGG TTA CCC CTC CCA TTA GCT AGA GCA GCC          2066
Pro Phe Leu Arg Leu Pro Arg Leu Pro Leu Pro Leu Ala Arg Ala Ala
        670                 675                 680

AAA CCT CCT CCA AAA CCT AGA TCC TCA CAA CCC TCT ACT GTG GAC TTG          2114
Lys Pro Pro Pro Lys Pro Arg Ser Ser Gln Pro Ser Thr Val Asp Leu
685                 690                 695

GCT TCT AAA GTT AAA CTA CCT GAA GGT CAT TAC CGG GTA GGG TCT CGA          2162
Ala Ser Lys Val Lys Leu Pro Glu Gly His Tyr Arg Val Gly Ser Arg
700                 705                 710                 715

GCC ATC TAC ACG TGC GAG TCG AGA TAC TAC GAA CTA CTT GGA TCT CAA          2210
Ala Ile Tyr Thr Cys Glu Ser Arg Tyr Tyr Glu Leu Leu Gly Ser Gln
                720                 725                 730

GGC AGA AGA TGT GAC TCT AAT GGA AAC TGG AGT GGT CGG CCA GCG AGC          2258
Gly Arg Arg Cys Asp Ser Asn Gly Asn Trp Ser Gly Arg Pro Ala Ser
            735                 740                 745

TGT ATT CCA GTT TGT GGA CGG TCA GAC TCT CCT CGT TCT CCT TTT ATC          2306
Cys Ile Pro Val Cys Gly Arg Ser Asp Ser Pro Arg Ser Pro Phe Ile
        750                 755                 760

TGG AAT GGG AAT TCT ACA GAA ATA GGT CAG TGG CCG TGG CAG GCA GGA          2354
Trp Asn Gly Asn Ser Thr Glu Ile Gly Gln Trp Pro Trp Gln Ala Gly
765                 770                 775

ATC TCT AGA TGG CTT GCA GAC CAC AAT ATG TGG TTT CTC CAG TGT GGA          2402
Ile Ser Arg Trp Leu Ala Asp His Asn Met Trp Phe Leu Gln Cys Gly
780                 785                 790                 795

GGA TCT CTA TTG AAT GAG AAA TGG ATC GTC ACT GCT GCC CAC TGT GTC          2450
Gly Ser Leu Leu Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val
                800                 805                 810

ACC TAC TCT GCT ACT GCT GAG ATT ATT GAC CCC AAT CAG TTT AAA ATG          2498
Thr Tyr Ser Ala Thr Ala Glu Ile Ile Asp Pro Asn Gln Phe Lys Met
            815                 820                 825

TAT CTG GGC AAG TAC TAC CGT GAT GAC AGT AGA GAC GAT GAC GAT TAT GTA          2546
Tyr Leu Gly Lys Tyr Tyr Arg Asp Asp Ser Arg Asp Asp Asp Asp Tyr Val
        830                 835                 840

CAA GTA AGA GAG GCT CTT GAG ATC CAC GTG AAT CCT AAC TAC GAC CCC          2594
Gln Val Arg Glu Ala Leu Glu Ile His Val Asn Pro Asn Tyr Asp Pro
845                 850                 855

GGC AAT CTC AAC TTT GAC ATA GCC CTA ATT CAA CTG AAA ACT CCT GTT          2642
Gly Asn Leu Asn Phe Asp Ile Ala Leu Ile Gln Leu Lys Thr Pro Val
860                 865                 870                 875

ACT TTG ACA ACA CGA GTC CAA CCA ATC TGT CTG CCT ACT GAC ATC ACA          2690
Thr Leu Thr Thr Arg Val Gln Pro Ile Cys Leu Pro Thr Asp Ile Thr
                880                 885                 890
```

-continued

```
ACA AGA GAA CAC TTG AAG GAG GGA ACA TTA GCA GTG GTG ACA GGT TGG        2738
Thr Arg Glu His Leu Lys Glu Gly Thr Leu Ala Val Val Thr Gly Trp
            895                 900                 905

GGT TTG AAT GAA AAC AAC ACC TAT TCA GAG ACG ATT CAA CAA GCT GTG        2786
Gly Leu Asn Glu Asn Asn Thr Tyr Ser Glu Thr Ile Gln Gln Ala Val
        910                 915                 920

CTA CCT GTT GTT GCA GCC AGC ACC TGT GAA GAG GGG TAC AAG GAA GCA        2834
Leu Pro Val Val Ala Ala Ser Thr Cys Glu Glu Gly Tyr Lys Glu Ala
    925                 930                 935

GAC TTA CCA CTG ACA GTA ACA GAG AAC ATG TTC TGT GCA GGT TAC AAG        2882
Asp Leu Pro Leu Thr Val Thr Glu Asn Met Phe Cys Ala Gly Tyr Lys
940                 945                 950                 955

AAG GGA CGT TAT GAT GCC TGC AGT GGG GAC AGT GGA GGA CCT TTA GTG        2930
Lys Gly Arg Tyr Asp Ala Cys Ser Gly Asp Ser Gly Gly Pro Leu Val
                960                 965                 970

TTT GCT GAT GAT TCC CGT ACC GAA AGG CGG TGG GTC TTG GAA GGG ATT        2978
Phe Ala Asp Asp Ser Arg Thr Glu Arg Arg Trp Val Leu Glu Gly Ile
            975                 980                 985

GTC AGC TGG GGC AGT CCC AGT GGA TGT GGC AAG GCG AAC CAG TAC GGG        3026
Val Ser Trp Gly Ser Pro Ser Gly Cys Gly Lys Ala Asn Gln Tyr Gly
        990                 995                1000

GGC TTC ACT AAA GTT AAC GTT TTC CTG TCA TGG ATT AGG CAG TTC ATT        3074
Gly Phe Thr Lys Val Asn Val Phe Leu Ser Trp Ile Arg Gln Phe Ile
    1005                1010                1015

TGAAACTGAT CTAAATATTT TAAGCATGGT TATAAACGTC TTGTTTCCTA TTATTGCTTT      3134

ACTAGTTTAA CCCATAAGAA GGTTAACTGG GTAAGGCACA AGGATCATTG TTTCTGTTTG      3194

TTTTTACAAA TGGTTATTTT AGTCAGTGAA TGAGAATAGT ATCCATTGAA GACTGTTACC      3254

TTTTATTCTA CCTTTTTATA TTACTATGTA AGTATTTGGG ATATCTTCTA CACATGAAAA      3314

TTCTGTCATT TTACCATAAA TTTGGTTTCT GGTGTGTGCT AAGTCCACCA GTAGAGAACG      3374

ATGTAATTTT CACTAGCACA TGAAATAAAT ATAGAACAAA TCTATTATAA ACTACCTTAA      3434

AAAAAAAAAA AAAA                                                        3448
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1019 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Val Leu Ala Ser Phe Leu Val Ser Gly Leu Val Leu Gly Leu Leu
 1               5                  10                  15

Ala Gln Lys Met Arg Pro Val Gln Ser Lys Gly Val Asp Leu Gly Leu
            20                  25                  30

Cys Asp Glu Thr Arg Phe Glu Cys Lys Cys Gly Asp Pro Gly Tyr Val
        35                  40                  45

Phe Asn Ile Pro Val Lys Gln Cys Thr Tyr Phe Tyr Arg Trp Arg Pro
    50                  55                  60

Tyr Cys Lys Pro Cys Asp Asp Leu Glu Ala Lys Asp Ile Cys Pro Lys
65                  70                  75                  80

Tyr Lys Arg Cys Gln Glu Cys Lys Ala Gly Leu Asp Ser Cys Val Thr
                85                  90                  95

Cys Pro Pro Asn Lys Tyr Gly Thr Trp Cys Ser Gly Glu Cys Gln Cys
            100                 105                 110

Lys Asn Gly Gly Ile Cys Asp Gln Arg Thr Gly Ala Cys Ala Cys Arg
```

-continued

```
                115                 120                 125

Asp Arg Tyr Glu Gly Val His Cys Glu Ile Leu Lys Gly Cys Pro Leu
     130                 135                 140

Leu Pro Ser Asp Ser Gln Val Gln Glu Val Arg Asn Pro Pro Asp Asn
145                 150                 155                 160

Pro Gln Thr Ile Asp Tyr Ser Cys Ser Pro Gly Phe Lys Leu Lys Gly
                165                 170                 175

Met Ala Arg Ile Ser Cys Leu Pro Asn Gly Gln Trp Ser Asn Phe Pro
            180                 185                 190

Pro Lys Cys Ile Arg Glu Cys Ala Met Val Ser Ser Pro Glu His Gly
        195                 200                 205

Lys Val Asn Ala Leu Ser Gly Asp Met Ile Glu Gly Ala Thr Leu Arg
    210                 215                 220

Phe Ser Cys Asp Ser Pro Tyr Tyr Leu Ile Gly Gln Glu Thr Leu Thr
225                 230                 235                 240

Cys Gln Gly Asn Gly Gln Trp Asn Gly Gln Ile Pro Gln Cys Lys Asn
                245                 250                 255

Leu Val Phe Cys Pro Asp Leu Asp Pro Val Asn His Ala Glu His Lys
            260                 265                 270

Val Lys Ile Gly Val Glu Gln Lys Tyr Gly Gln Phe Pro Gln Gly Thr
        275                 280                 285

Glu Val Thr Tyr Thr Cys Ser Gly Asn Tyr Phe Leu Met Gly Phe Asp
    290                 295                 300

Thr Leu Lys Cys Asn Pro Asp Gly Ser Trp Ser Gly Ser Gln Pro Ser
305                 310                 315                 320

Cys Val Lys Val Ala Asp Arg Glu Val Asp Cys Asp Ser Lys Ala Val
                325                 330                 335

Asp Phe Leu Asp Asp Val Gly Glu Pro Val Arg Ile His Cys Pro Ala
            340                 345                 350

Gly Cys Ser Leu Thr Ala Gly Thr Val Trp Gly Thr Ala Ile Tyr His
        355                 360                 365

Glu Leu Ser Ser Val Cys Arg Ala Ala Ile His Ala Gly Lys Leu Pro
    370                 375                 380

Asn Ser Gly Gly Ala Val His Val Val Asn Asn Gly Pro Tyr Ser Asp
385                 390                 395                 400

Phe Leu Gly Ser Asp Leu Asn Gly Ile Lys Ser Glu Glu Leu Lys Ser
                405                 410                 415

Leu Ala Arg Ser Phe Arg Phe Asp Tyr Val Arg Ser Ser Thr Ala Gly
            420                 425                 430

Lys Ser Gly Cys Pro Asp Gly Trp Phe Glu Val Asp Glu Asn Cys Val
        435                 440                 445

Tyr Val Thr Ser Lys Gln Arg Ala Trp Glu Arg Ala Gln Gly Val Cys
    450                 455                 460

Thr Asn Met Ala Ala Arg Leu Ala Val Leu Asp Lys Asp Val Ile Pro
465                 470                 475                 480

Asn Ser Leu Thr Glu Thr Leu Arg Gly Lys Gly Leu Thr Thr Thr Trp
                485                 490                 495

Ile Gly Leu His Arg Leu Asp Ala Glu Lys Pro Phe Ile Trp Glu Leu
            500                 505                 510

Met Asp Arg Ser Asn Val Val Leu Asn Asp Asn Leu Thr Phe Trp Ala
        515                 520                 525

Ser Gly Glu Pro Gly Asn Glu Thr Asn Cys Val Tyr Met Asp Ile Gln
    530                 535                 540
```

-continued

Asp Gln Leu Gln Ser Val Trp Lys Thr Lys Ser Cys Phe Gln Pro Ser
545 550 555 560

Ser Phe Ala Cys Met Met Asp Leu Ser Asp Arg Asn Lys Ala Lys Cys
565 570 575

Asp Asp Pro Gly Ser Leu Glu Asn Gly His Ala Thr Leu His Gly Gln
580 585 590

Ser Ile Asp Gly Phe Tyr Ala Gly Ser Ser Ile Arg Tyr Ser Cys Glu
595 600 605

Val Leu His Tyr Leu Ser Gly Thr Glu Thr Val Thr Cys Thr Thr Asn
610 615 620

Gly Thr Trp Ser Ala Pro Lys Pro Arg Cys Ile Lys Val Ile Thr Cys
625 630 635 640

Gln Asn Pro Pro Val Pro Ser Tyr Gly Ser Val Glu Ile Lys Pro Pro
645 650 655

Ser Arg Thr Asn Ser Ile Ser Arg Val Gly Ser Pro Phe Leu Arg Leu
660 665 670

Pro Arg Leu Pro Leu Pro Leu Ala Arg Ala Ala Lys Pro Pro Pro Lys
675 680 685

Pro Arg Ser Ser Gln Pro Ser Thr Val Asp Leu Ala Ser Lys Val Lys
690 695 700

Leu Pro Glu Gly His Tyr Arg Val Gly Ser Arg Ala Ile Tyr Thr Cys
705 710 715 720

Glu Ser Arg Tyr Tyr Glu Leu Leu Gly Ser Gln Gly Arg Arg Cys Asp
725 730 735

Ser Asn Gly Asn Trp Ser Gly Arg Pro Ala Ser Cys Ile Pro Val Cys
740 745 750

Gly Arg Ser Asp Ser Pro Arg Ser Pro Phe Ile Trp Asn Gly Asn Ser
755 760 765

Thr Glu Ile Gly Gln Trp Pro Trp Gln Ala Gly Ile Ser Arg Trp Leu
770 775 780

Ala Asp His Asn Met Trp Phe Leu Gln Cys Gly Gly Ser Leu Leu Asn
785 790 795 800

Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Thr Tyr Ser Ala Thr
805 810 815

Ala Glu Ile Ile Asp Pro Asn Gln Phe Lys Met Tyr Leu Gly Lys Tyr
820 825 830

Tyr Arg Asp Asp Ser Arg Asp Asp Asp Tyr Val Gln Val Arg Glu Ala
835 840 845

Leu Glu Ile His Val Asn Pro Asn Tyr Asp Pro Gly Asn Leu Asn Phe
850 855 860

Asp Ile Ala Leu Ile Gln Leu Lys Thr Pro Val Thr Leu Thr Thr Arg
865 870 875 880

Val Gln Pro Ile Cys Leu Pro Thr Asp Ile Thr Thr Arg Glu His Leu
885 890 895

Lys Glu Gly Thr Leu Ala Val Val Thr Gly Trp Gly Leu Asn Glu Asn
900 905 910

Asn Thr Tyr Ser Glu Thr Ile Gln Gln Ala Val Leu Pro Val Val Ala
915 920 925

Ala Ser Thr Cys Glu Glu Gly Tyr Lys Glu Ala Asp Leu Pro Leu Thr
930 935 940

Val Thr Glu Asn Met Phe Cys Ala Gly Tyr Lys Lys Gly Arg Tyr Asp
945 950 955 960

Ala Cys Ser Gly Asp Ser Gly Gly Pro Leu Val Phe Ala Asp Asp Ser
965 970 975

-continued

```
Arg Thr Glu Arg Arg Trp Val Leu Glu Gly Ile Val Ser Trp Gly Ser
            980                 985                 990

Pro Ser Gly Cys Gly Lys Ala Asn Gln Tyr Gly Gly Phe Thr Lys Val
        995                1000                1005

Asn Val Phe Leu Ser Trp Ile Arg Gln Phe Ile
    1010                1015
```

We claim:

1. An isolated DNA molecule encoding a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4.

2. The DNA molecule of claim **3**, having a nucleotide sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3.

3. The DNA molecule of claim **2**, wherein said DNA molecule has the nucleotide sequence of SEQ ID NO:1, wherein nucleotides 1 to 568 have been deleted.

4. A recombinant vector comprising DNA encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:4.

5. The recombinant vector of claim **4**, wherein said DNA has a nucleotide sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3.

6. A host cell transformed with said recombinant vector of claim **5**.

7. The host cell of claim **6**, wherein said host cell is selected from the group consisting of *E. coli*, a mammalian cell, an insect cell, and a yeast.

8. A method for producing a protein having the enzymatic activity of Factor C protein in assays for Gram negative bacterial endotoxin, comprising culturing said host cell according to claim **6** under conditions such that said DNA molecule is expressed, and recovering said protein.

9. The recombinant vector of claim **4**, wherein the vector portion is a member selected from the group consisting of λgt 22, pGEM 11Zf(+), pGEMEX-1, pET 3b, YEpsec 1, pEMBLyex 4, pIC 9 and pHIL D2.

10. A host cell transformed with said recombinant vector of claim **4**.

11. The host cell of claim **10**, wherein said host cell is selected from the group consisting of *E. coli*, a mammalian cell, an insect cell, and a yeast.

12. A method of producing a protein having the same enzymatic activity as Factor C protein in assays for Gram negative bacterial endotoxin, comprising culturing said host cell of claim **10** under conditions in which said DNA molecule is expressed, and recovering said protein.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,716,834
DATED : Feb. 10, 1998
INVENTOR(S) : DING et al.

It is certified that error appears in the above-indentified patent and that said Letters Patent is hereby corrected as shown below:

On the Cover Page

[54] In the Title

change "CDNA" to --cDNA--

Column 1, line 1

change "CDNA" to --cDNA--

Column 2, line 51

change “pIC” to --pPIC--

Column 4, line 20

change “sublconed” to --subcloned--

Column 8, line 13

Change “Na2EDTA” to --Na_2EDTA--

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,716,834
DATED : Feb. 10, 1998
INVENTOR(S) : DING et al.

It is certified that error appears in the above-indentified patent and that said Letters Patent is hereby corrected as shown below:

Column 13, line 27
change "C. rotundicauda" to
--*C. rotundicauda*--

Column 54, line 4 of claim 9
change "pIC" to --pPIC--

Signed and Sealed this
Twenty-second Day of September, 1998

Attest:

BRUCE LEHMAN

Attesting Officer *Commissioner of Patents and Trademarks*

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,716,834

DATED : February 10, 1998

INVENTOR(S) : DING et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 53:

Claim 2, change "claim 3" to --claim 1--

Signed and Sealed this

Fourth Day of January, 2000

Attest:

Attesting Officer

Acting Commissioner of Patents and Trademarks

Molecular Marine Biology and Biotechnology (1995) 4(1), 90–103

Molecular cloning and sequence analysis of Factor C cDNA from the Singapore horseshoe crab, *Carcinoscorpius rotundicauda*

Jeak L. Ding* and M. Anthony A. Navas III†
Marine Biotechnology Laboratory, and BioScience Centre, Department of Zoology, National University of Singapore, Kent Ridge, Singapore 0511

Bow Ho
Department of Microbiology, National University of Singapore, Kent Ridge, Singapore 0511

Abstract

Two forms of Factor C cDNAs: CrFC21 (3448 bp) and CrFC26 (4182 bp) have been cloned into λgt22. CrFC26 includes 568 nucleotides of 5′ untranslated region (5′ UTR) containing seven ATGs before the real initiation site, an open reading frame (ORF) of 3249 nucleotides, a stop codon, and 365 nucleotides of 3′ untranslated sequence. There are four polyadenylation signals and six potential glycosylation sites. The ORF codes for a signal peptide of 24 amino acids and a Factor C zymogen of 1059 residues. The CrFC21 lacks most of the 5′ UTR, and has some base changes in its ORF. The predicted secondary mRNA structures of the 5′ end of CrFC26 showed numerous stem-and-loop structures, thus obscuring its real start codon. In contrast, CrFC21 has a well-exposed AUG start site, and expresses Factor C in transcription-translation reactions in vitro. There is a typical serine protease catalytic triad of Asp-His-Ser, which is structurally like prothrombin, but catalytically more similar to trypsin. Although an overall homology of 97.7% was observed in comparison with the *Tachypleus tridentatus* Factor C (TtFC) cDNA, there were notable differences in the restriction sites and subtle base substitutions in the CrFC cDNA. The high degree of homology between Factor C from *T. tridentatus* and *C. rotundicauda* substantiates, at the molecular level, the proximity of these two species in the course of evolution. This finding contravenes the apparent disparities with respect to their morphology, ecological habitat, and taxonomical classification.

*Correspondence should be sent to this author.
†Present address: Department of Biochemistry, Baylor College, Houston, Texas, USA.

Introduction

There are four extant species of horseshoe crabs: *Limulus polyphemus, Tachypleus tridentatus, Tachypleus gigas,* and *Carcinoscorpius rotundicauda.* Within the predominant type of blood cell, amoebocytes, are enzymes belonging to the family of serine proteases that sequentially undergo limited proteolysis (Iwanaga et al., 1985) upon endotoxin activation to finally convert coagulogen into an insoluble gel, coagulin (Solum, 1973). This cascade of enzymes and coagulogen constitutes the commercially available amoebocyte lysate used for detecting endotoxin. The latter is extremely ubiquitous and indomitable, contaminating parenteral preparations, water, food, and pharmaceutical products. The *Limulus* amoebocyte lysate (LAL) is a rapid and sensitive in vitro assay for detecting femtogram levels of endotoxin (Ho, 1983).

The elucidation of the molecular events that are responsible for the subsequent gelation of the amoebocyte lysate was attributed to the isolation and characterization of the serine proteases from either *L. polyphemus* or *T. tridentatus*. These include Factor C (Nakamura et al., 1986) and proclotting enzyme (Nakamura et al., 1985). However, to date, there is a lack of information on similar enzymes from their mud-dwelling Southeast Asian counterpart, *C. rotundicauda,* except our recent reports on purification and characterization of its Factor C enzyme (Navas et al., 1990; Ding et al., 1993a; Ho et al., 1993).

Factor C is the endotoxin-sensitive intracellular serine protease zymogen that initiates the coagulation cascade system. It exists as single- and double-chain forms in *C. rotundicuada* (Ding et al., 1993a). While the catalytic site of the molecule is found in the light chain, the endotoxin-binding site is located at the NH_2-terminal heavy chain region (Nakamura et al., 1988). Muta et al. (1991) have cloned the *T. tridentatus* Factor C, albeit in two partial overlapping fragments: pFC 41 and pFC 53, henceforth re-

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Seikagaku Corporation Exhibit 2004 - Page 297 of 310
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

ferred to as TtFC41 and TtFC53, respectively. These two fragments gave a composite sequence of 3474 bp. We report, here, directional cloning of the full-length *C. rotundicauda* Factor C (CrFC) cDNA in λgt22 vector, its characterization by restriction mapping and DNA sequencing, homology comparison studies with TtFC cDNAs and other serine proteases in the EMBL gene bank, the computational analysis of the 5′ region of CrFC cDNAs and their corresponding predicted mRNA secondary structures, and the evolutionary significance of CrFC. The commercial significance of recombinant Factor C is also implied.

Results and Discussion

Northern analysis of C. rotundicauda *amoebocyte RNA*

Northern hybridization of *C. rotundicauda* RNA using TtFC probe under low stringency of 42°C showed three bands corresponding to 15.8, 5.9, and 4.0 kNt (Figure 1a). However, subsequent hybridization under high stringency with homologous EE fragment of CrFC26 revealed only the 4.0 kNt RNA band (Figure 1b). The other forms of CrFC mRNAs could represent transcripts resulting from alternative splicing as indicated by the presence of alternative polyadenylation signals (see Fig. 5). Since total RNA was used in the earlier blot, the 15.8-kNt band could have been the hnRNA transcript.

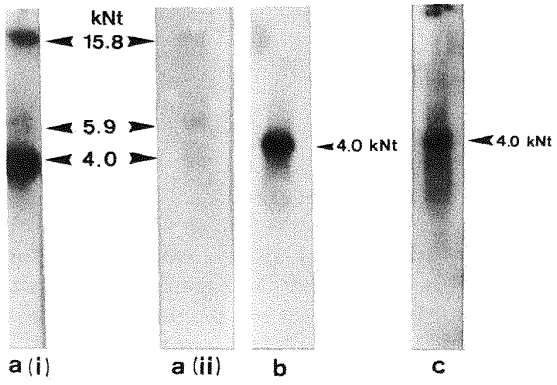


Figure 1. Northern blot analysis using TtFC53 as probe on 10 μg total RNA in lane a (i) and 1 μg of mRNA in lane a (ii); pCrFC26 (*Eco*RI–*Eco*RI fragment) as probe on 10 μg of total RNA in lane b; and the 368-bp *Eco*RI–*Nde*I (5′ end) of pCrFC26 as probe on 10 μg of total RNA in lane c. A single band of approximately 4 kNt in both blots (b) and (c) indicates that the cDNA isolated is full-length and the entire 5′ end of crFC26 unequivocally belongs to this species of Factor C.

The Factor C cDNA λ recombinant clones

From 250,000 recombinant clones, only 0.03% of the cDNA library was positive for Factor C cDNA, indicating its status as a rare copy gene. Following secondary and tertiary screenings, 16 λ recombinant clones, henceforth referred to as CrFC cDNAs, were further characterized. Their orientations were determined, and initial restriction maps were analyzed for *Eco*RI and *Not*I sites (Figure 2).

Double digestion of the CrFC clones with *Eco*RI and *Not*I followed by Southern blot hybridization with the *T. tridentatus* Factor C (TtFC53) cDNA probe showed positive signals for all the CrFC clones (data not shown). When the same membrane was deprobed and reprobed with the 3′-end EN fragment of CrFC26, or the whole of TtFC41, two of the CrFC clones, CrFC45 and CrFC63, did not

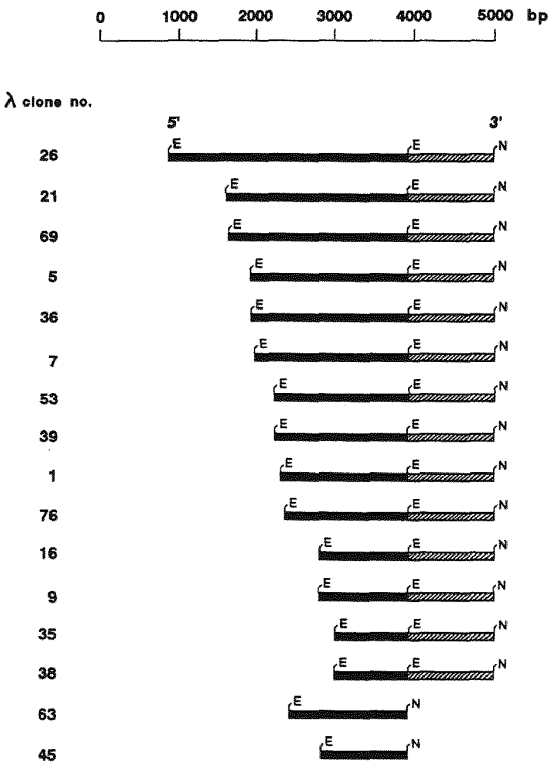


Figure 2. The initial restriction map of the various Factor C cDNAs isolated and purified from the *C. rotundicauda* amoebocyte cDNA library based on Southern analysis with TtFC53 and the 1100-bp *Eco*RI–*Not*I fragment. The cDNAs are arranged in decreasing order of size and are labeled with the restriction sites *Eco*RI (E) and *Not*I (N). Fragments homologus to TtFC53 are drawn with solid boxes while those homologous to the E–N fragment are drawn with hatched boxes.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

cross-hybridize with either probe, suggesting that these two clones are not similar to the other CrFC cDNA clones. They do not have an internal *Eco*RI site, but are homologous to only the 5′-end region (EE fragment) of other CrFC clones and the TtFC53.

Comparison of the restriction maps of CrFC and TtFC cDNAs

The complete restriction maps of the two longest clones of *C. rotundicauda* Factor C cDNAs, CrFC21 and CrFC26, are shown in Figure 3, in comparison with the maps of the two overlapping clones of *T. tridentatus* Factor C, TtFC53 and TtFC41 (Muta et al., 1991). CrFC26 has a longer 5′ terminus compared with the TtFC53. Based on the restriction maps of TtFC53 and TtFC41 clones, the cDNAs from both species have the same locations for restriction sites *Sty*I, *Sal*I, and *Sac*I. However, notable differences were also observed between the two genes: CrFC has two *Bam*HI sites compared with only one in TtFC53. On the other hand, the former only has two *Hind*III sites compared with three in the latter.

The TtFC41, on the other hand, encompasses the 3′ region of the TtFC. It extends from the *SstI*/*Sac*I site up to the 3′ end, and it was found to contain the internal *Eco*RI site relative to that of CrFC. The other common site between TtFC41 and CrFC is the *Pst*I site proximal to the 3′ end, but the second *Pst*I site near the 5′ end of TtFC41 was absent in the CrFC cDNAs. The *Bam*HI and *Hind*III sites in TtFC41 were absent in the CrFC26.

The DNA sequence and derived amino acid sequence of CrFC cDNA

The DNA sequences of CrFC21 and CrFC26 were determined on both strands by the Sanger dideoxy sequencing method. Figure 4 shows the sequencing

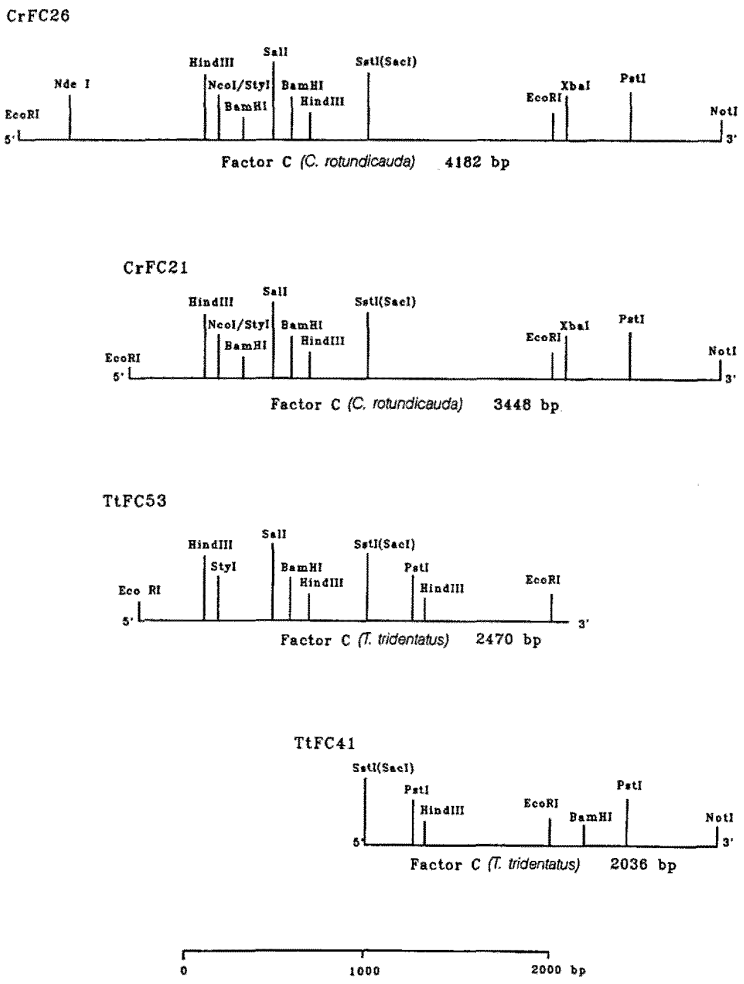


Figure 3. A comparison between the complete restriction maps of CrFC26, CrFC21 *(C. rotundicauda),* and TtFC53 and TtFC41 (*T. tridentatus,* adapted from Muta et al., 1991).

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

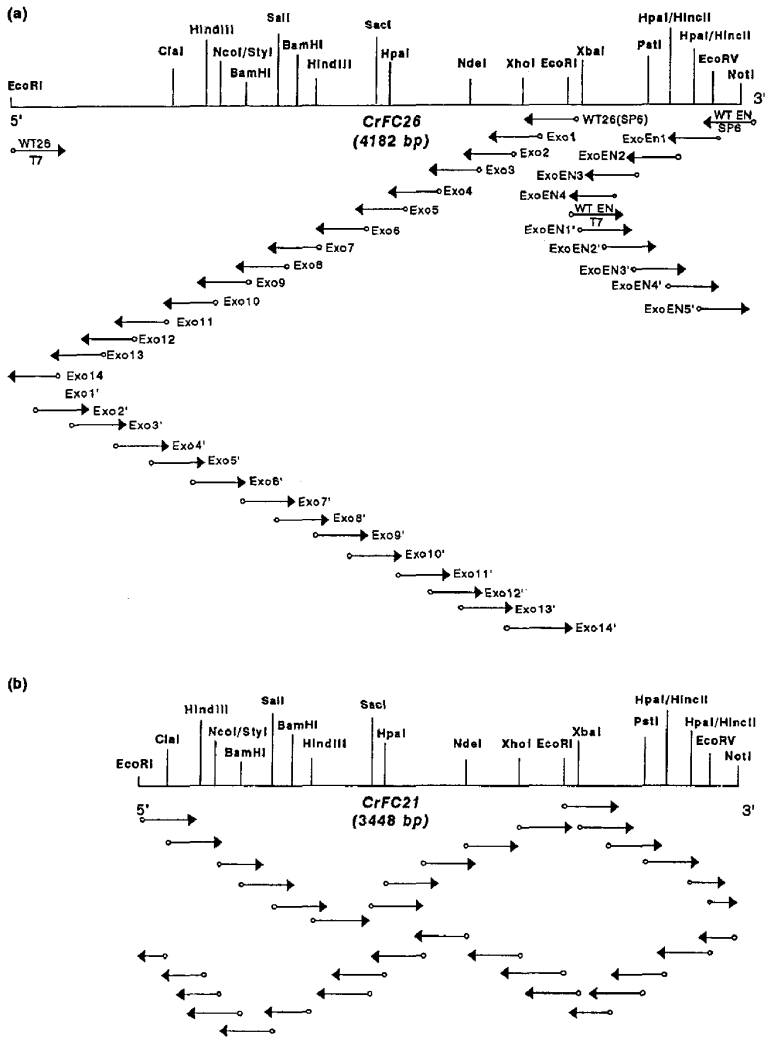


Figure 4. Sequencing strategies of (a) CrFC26 and (b) CrFC21. Deletion subclones were prepared from pEE and pEN in both directions. The arrows indicate the direction and the extent of the sequences obtained. T7 promoter primer (o⟶); SP6 promoter primer (⟵o).

strategies. The CrFC26 was found to be 4182 bp long (Figure 5). This is consistent with the Northern analysis of the amoebocyte total RNA with homologous EE insert, which showed only one hybridizing band at approximately 4 kNt (Figure 1b). This complete cDNA includes 568 nucleotides of 5′ untranslated sequence containing seven ATGs before the real initiation site, an ORF of 3249 nucleotides coding for a protein of 1083 amino acids, a stop codon, and 365 nucleotides of 3′ untranslated sequence. At the 3′ end of the cDNA, the canonical hexanucleotide sequence AATAAA is present 19 nucleotides upstream from the polyadenylation site (at nucleotide position 4142). There are three other AATAAA sequences found at nucleotide positions 183, 239, and 2474, respectively. The ORF codes for a signal peptide of 24 amino acids and a Factor C zymogen of 1059 residues, which, prior to *N*-glycosylation, has a calculated molecular weight of 120,244 daltons. This is close to the estimated molecular weight of single- and double-chain Factor C enzymes purified from the amoebocytes (Ding et al., 1993a). The CrFC21 cDNA of 3448 nucleotides in length, its structural features, and its derived amino acid sequence are shown in Figure 5. With the exception of the absence of the majority of 5′ UTR and some subtle base changes, the structural features within the ORF of CrFC21 bear close resemblance to those of CrFC26.

The unique proteolytic site due to endotoxin-activation of the Factor C enzyme is found between Phe-Ile: F^{762}-I^{763} for CrFC21, and F^{826}-I^{827} for CrFC26, thus indicating the integrity of the cloned Factor C cDNA, and its potentials in endotoxin detection.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

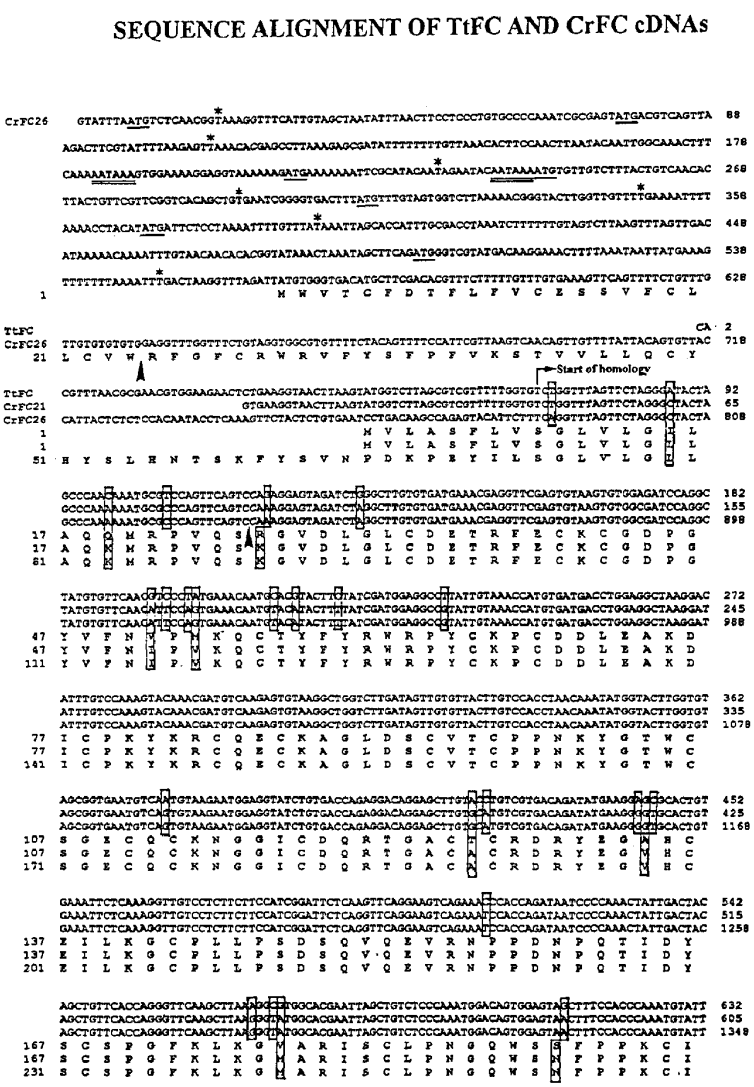


Figure 5. Alignment of the complete DNA sequence and deduced amino acid sequence of *C. rotundicauda* Factor C (CrFC26 and CrFC21) with the Factor C of *T. tridentatus* (TtFC, upper sequence is adapted from Muta et al. 1991). The numbering of the amino acid residues is found at the left while that of the DNA (in bp) is found at the right. The site of truncation of the signal peptide is indicated by a small arrowhead (▲). The "start of homology" region is indicated by a directional arrow (↱). The differences in both the DNA and amino acid sequences are boxed (□). The cleavage site of the Factor C protein into heavy- and light-chain intermediates is indicated by a hollow thick arrow (⇦) between residues R and S, while proteolysis of the light chain into A and B due to endotoxin-activation is indicated by a solid thick arrow (⬅) between residues F and I. The potential glycosylation sites are marked with closed diamonds (♦). The amino acid residues constituting the catalytic triad by analogy with trypsin are indicated by white stars (✪). The four AATAAA sites are marked with double underlines (═). The seven false start ATG sites are indicated with single underlines. These sites are terminated shortly by in-frame stop codons indicated by black stars (⋆) located several bases downstream.

Homology between the DNA sequences of CrFC and TtFC

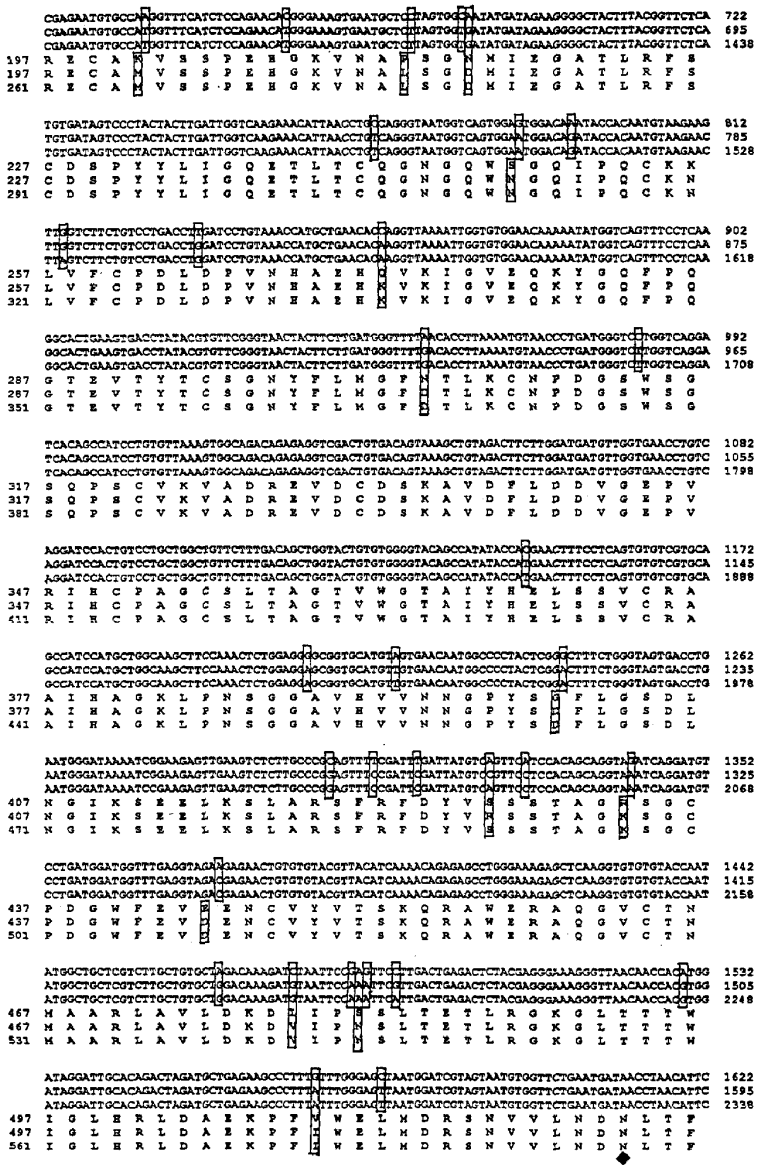
In comparing CrFC26 and TtFC53 (Figure 5), the *C. rotundicauda* cDNA was found to be longer by 716 nucleotides at the 5′ end, and it also has an extra 64 amino acids upstream from the starting *met*. The start of homology of both the DNA and amino acid sequences between the two species was found at nucleotide position 785 for CrFC26 and at position 69 for TtFC. The homologous region extends through to the poly(A) tail of both genes, giving an overall homology of 97.7%. Most of the dissimilarities, however, were found in the 3′ UTR, where a lower percentage homology of 86% was observed.

On the other hand, similar to TtFC, there are six potential *N*-glycosylation sites within the CrFC sequence Asn-Xaa-Ser/Thr.

A majority of the differences between the two sequences found within the ORF were due to base substitutions (Figure 5). In fact, these substitutions explain the perceived differences in the restriction sites previously obtained by restriction mapping (see Figure 3). For instance, the absence of the *Bam*HI (GGATCC) site in TtFC, which in CrFC26 is located 1549 bp from the 5′ end, was due to the substitution of the first G with a T in TtFC. The converse is shown with the *Bam*HI site at 2430 of TtFC in which the second C is substituted with a T in CrFC. Similarly, the *Pst*I site (CTGCAG) of TtFC

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Figure 5. *Continued*



(at 1683 bp from its 5′ end) was not found in CrFC because the first C was substituted with a T. Again, a C-to-T base substitution (at 1700 and 2443 bp positions for CrFC21 and CrFC26, respectively) was responsible for the loss of the third *Hind*III (AAGCTT) in CrFC, which in TtFC was located 1724 bp from the 5′ end. A change from A (at 646 of TtFC) to T (at 1362 of CrFC26) gave rise to either an *Sty*I or *Nco*I site in CrFC26, whereas TtFC is recognized by only *Sty*I. Base substitution was also shown in the *Xba*I (TCTAGA) site at 3101 bp of CrFC, where the first A is substituted with a C in TtFC.

The 5′ noncoding region of CrFC and its computational analysis

Figure 6 shows the secondary structure predicted for the 5′ ends of the mRNA of CrFC21 and CrFC26. Numerous hairpin stems and loops are seen for the CrFC26, obscuring its real start codon, whereas CrFC21 showed a well-exposed AUG start site for ribosome binding.

Exceptionally long 5′ untranslated sequences occur on unusually interesting mRNAs such as epidermal growth factor (EGF) (Scott et al., 1983), oncogenes (Watt et al., 1983), and heat shock proteins

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

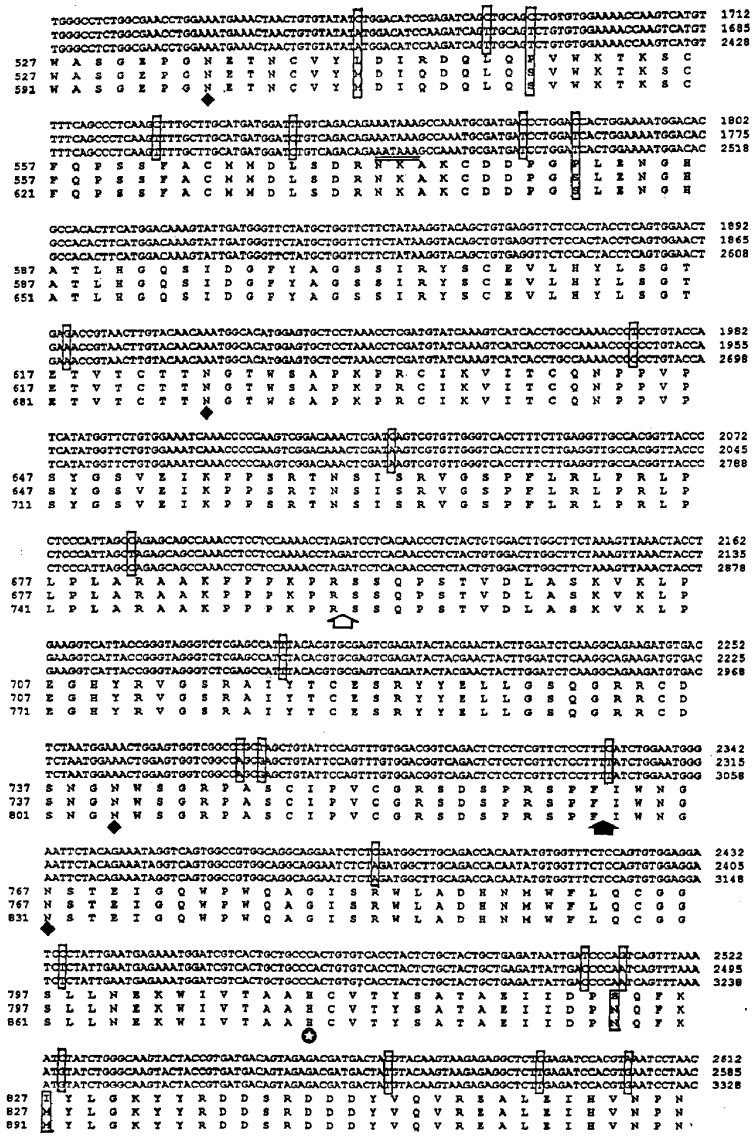


Figure 5. *Continued*

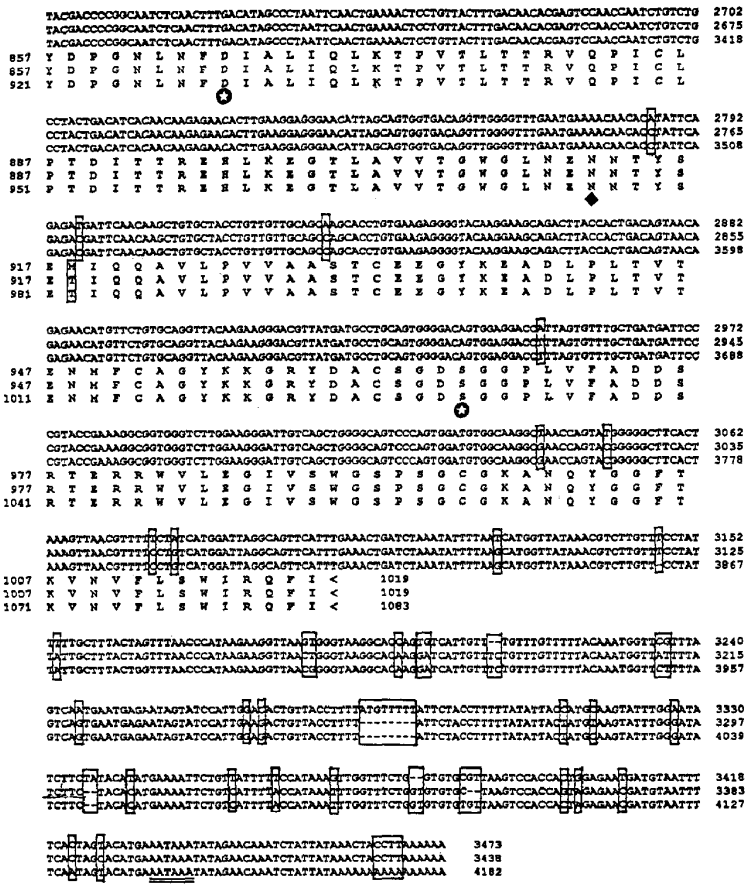
(Ingolia and Craig, 1981). This invites speculation that long structures of the 5′ noncoding region participate in the regulated expression of these genes, including that of Factor C. However, a long 5′ noncoding sequence in some mRNAs may have a deleterious effect (Kozak, 1983), especially if it contains secondary structures that may deflect ribosomes from the authentic initiation site that lies farther downstream. This was true for the Semliki Forest virus (Lehtovaara et al., 1982) and the recombinant fish antifreeze protein III gene (Li et al., 1991). This could also apply to the CrFC26 since there are seven false ATG sites (boxed in Figure 5) upstream from the real ATG site at nucleotide position 569. All these false start sites were followed shortly by in-frame stop sites downstream. This configuration agrees with the ribosome scanning-reinitiation model of translation in eukaryotic genes (Kozak, 1986). This is probably one reason why there exists another form of Factor C mRNA having a much shorter 5′ noncoding sequence as in CrFC21, where it may be translated more efficiently.

Variant forms of the CrFC gene

The Factor C cDNA sequence indicates the existence of at least two types of Factor C mRNA in *C. rotundicauda*. The 5′-end terminal sequences of CrFC26 and CrFC21 differed significantly. Compari-

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Figure 5. *Continued*

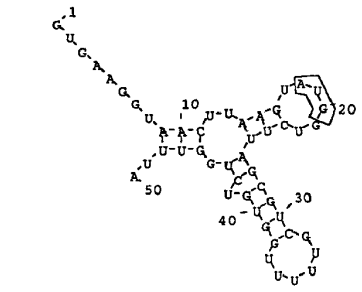


son of the sequences of *C. rotundicauda* Factor C clones CrFC21 and CrFC26 with that of *T. tridentatus* (TtFC53) shows that CrFC21 and TtFC53 share precise homology at their 5′ ends. Yet, the most notable difference in the DNA sequences of Factor C from the two species of horseshoe crabs is the existence of an extra 716 nucleotides at the 5′ end of the CrFC26 (Figure 7). Thus, CrFC21 was not merely a truncated species of CrFC26; rather, it was derived from a totally distinct species of mRNA that is exactly identical to the one that gave rise to the *T. tridentatus* Factor C cDNA. To demonstrate unequivocally the authenticity of the 5′ noncoding region of CrFC26, a further Northern blot analysis was performed, but using only the first 368 bp of CrFC26 (fragment flanked by *Eco*RI and *Nde*I) as probe. The Northern blot result in Figure 1c clearly demonstrates that the entire 5′ end of CrFC26 truly belongs to this particular species of Factor C mRNA. Thus, the existence of the two types of *C. rotundicauda* Factor C cDNA could be attributable to differential splicing of the initial primary transcript around its 5′ terminal.

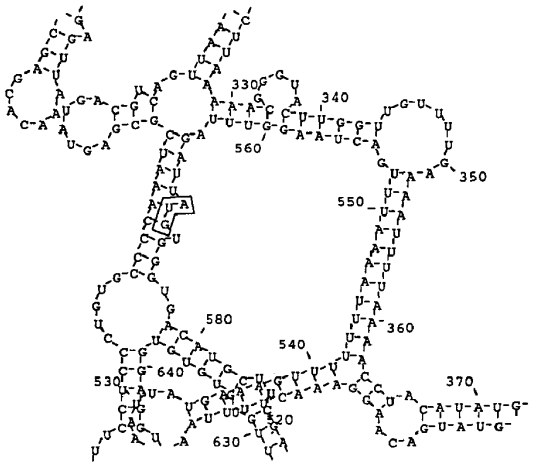
Structural domains and evolutionary significance of C. rotundicauda *Factor C*

The secondary folding structures of several blood clotting factors in the human coagulation system are usually identified by common domains derived from homologous regions in the DNA sequence (Furie and Furie, 1988). Typically, all these factors have a signal peptide at its amino-terminus, and a catalytic domain at the carboxy-terminus, which for serine proteases like *C. rotundicauda* Factor C, also contain the catalytic triad, His^{809}-Asp^{865}-Ser^{966} for CrFC21 and His^{873}-Asp^{929}-Ser^{1030} for CrFC26 (Figure 5). A hydropathy profile based on the formula of Kyte and Doolittle (1982) of the first 59 amino acids of CrFC26 showed the first 24 residues to be the putative signal peptide rich in hydrophobic amino acids (Figure 8). CrFC26 has five short consensus repeats (residues 206–259, 263–318,

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)



(a) CrFC 21



(b) CrFC 26

Figure 6. The computational predictions of the secondary structures of mRNAs of (a) CrFC21 and (b) CrFC26. The leader sequence of CrFC26 shows numerous hairpin stems and loops. The authentic AUG start codon is boxed. The start codon in CrFC21 is exposed in a loop, and appears more accessible.

324–385, 640–698, and 758–812), one lectin-domain (500–632), one EGF-like domain (167–200), a Cys-rich region (90–165), and one Pro-rich region (732–754). The serine protease domain is found at positions 827–1083. No propeptide sequence, kringle domains (Park and Tulinsky, 1986), amino acids stacks, or finger domains were found.

Scanning the EMBL gene bank in a homology search for the serine protease catalytic domain of CrFC revealed that it is structurally closest to *T. tridentatus* Factor C, with 97.7% homology in 777 overlapping nucleotides (Table 1).

From the overall comparison with serine proteases from other species, the CrFC is structurally closest to prothrombin and thrombin. This finding agrees well with the preference of Factor C for hydrolysis of synthetic substrates of thrombin. Surprisingly, a human liver hepsin (Leytus et al., 1986) exhibited a slightly higher level of homology with CrFC than did other blood coagulation factors. Although the function of hepsin is still unknown, this serine protease, which lacks a typical signal peptide but instead has a transmembrane domain, was nev-

Table 1. Percentage homology comparison of the catalytic domain of *Carcinoscorpius rotundicauda* Factor C with other serine proteases based on entries in the EMBL data bank.

Gene	Homology (%)	Overlap (bp)
Tachypleus tridentatus Factor C	97.7	777
Onchorhynchus mykiss thrombin	54.4	709
Rattus norvegicus thrombin	52.2	624
Rat prothrombin	52.2	624
Acipencer transmontanus thrombin	52.8	623
Oryctolagus cuniculus thrombin	57.3	595
Mouse prothrombin	52.7	509
Human blood coagulation Factor VII	55.7	420
Guinea pig Factor IX	57.4	296
Human liver hepsin	59.0	278

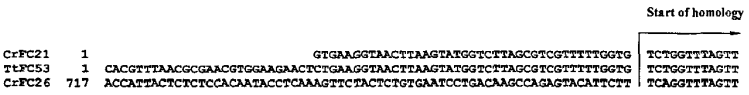
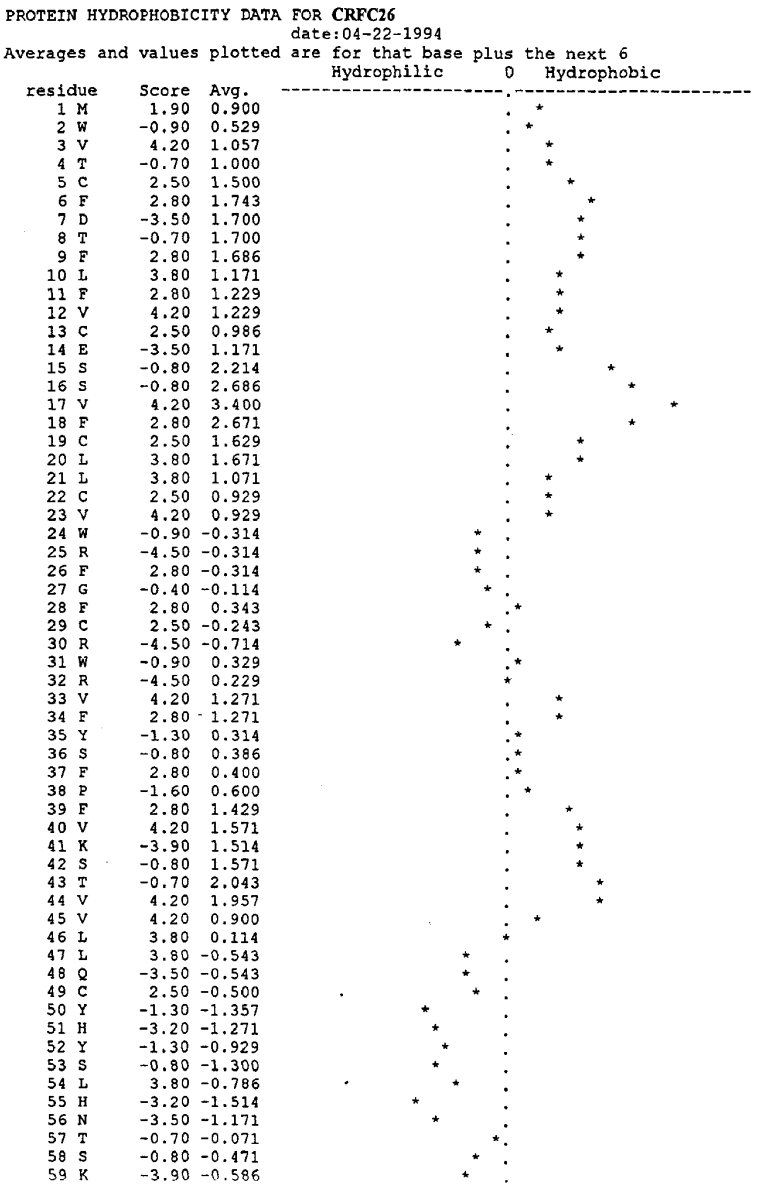


Figure 7. Comparison of the amino-terminal nucleotide sequences of different Factor C cDNAs isolated from the horseshoe crabs. The clones TtFC53 and TtFC41 is from *T. tridentatus* (Muta et al., 1991) while clones CrFC21 and CrFC26 are from *C. rotundicauda* (this work). CrFC21 and TtFC53 have identical 5′-end sequence, while CrFC26 has an extra 716 nucleotides at its 5′ end.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)



PROTEIN HYDROPHOBICITY DATA FOR **CRFC26**
date:04-22-1994
Averages and values plotted are for that base plus the next 6

residue		Score	Avg.
1	M	1.90	0.900
2	W	-0.90	0.529
3	V	4.20	1.057
4	T	-0.70	1.000
5	C	2.50	1.500
6	F	2.80	1.743
7	D	-3.50	1.700
8	T	-0.70	1.700
9	F	2.80	1.686
10	L	3.80	1.171
11	F	2.80	1.229
12	V	4.20	1.229
13	C	2.50	0.986
14	E	-3.50	1.171
15	S	-0.80	2.214
16	S	-0.80	2.686
17	V	4.20	3.400
18	F	2.80	2.671
19	C	2.50	1.629
20	L	3.80	1.671
21	L	3.80	1.071
22	C	2.50	0.929
23	V	4.20	0.929
24	W	-0.90	-0.314
25	R	-4.50	-0.314
26	F	2.80	-0.314
27	G	-0.40	-0.114
28	F	2.80	0.343
29	C	2.50	-0.243
30	R	-4.50	-0.714
31	W	-0.90	0.329
32	R	-4.50	0.229
33	V	4.20	1.271
34	F	2.80	1.271
35	Y	-1.30	0.314
36	S	-0.80	0.386
37	F	2.80	0.400
38	P	-1.60	0.600
39	F	2.80	1.429
40	V	4.20	1.571
41	K	-3.90	1.514
42	S	-0.80	1.571
43	T	-0.70	2.043
44	V	4.20	1.957
45	V	4.20	0.900
46	L	3.80	0.114
47	L	3.80	-0.543
48	Q	-3.50	-0.543
49	C	2.50	-0.500
50	Y	-1.30	-1.357
51	H	-3.20	-1.271
52	Y	-1.30	-0.929
53	S	-0.80	-1.300
54	L	3.80	-0.786
55	H	-3.20	-1.514
56	N	-3.50	-1.171
57	T	-0.70	-0.071
58	S	-0.80	-0.471
59	K	-3.90	-0.586

Figure 8. Hydropathy analysis of the first 59 deduced amino acid sequence of *C. rotundicauda* Factor C. The hydrophobic peak spanning residues 1–24 represents the putative signal sequence.

ertheless shown to have characteristics typical of trypsin. Based on the primary structure of CrFC, it was also found to the catalytically similar to trypsin-type serine protease, although structurally closer to thrombin. This may be explained by the Asp^{1024} (as opposed to Ser) in CrFC26, analogous to Asp^{189} in trypsin (Hartley, 1970), which strongly suggests that CrFC possesses a substrate specificity similar to that of trypsin.

The high degree of homology of the DNA sequences of CrFC and TtFC substantiates, at the molecular level, the proximity of these two species of horseshoe crab in the course of evolution. Calculations based on mutation distances among the amino acid sequences of coagulogen purified from the four extant species of horseshoe crab have indicated that *C. rotundicauda* could have diverged more recently *from T. tridentatus* than did *L. polyphemus* and *T. gigas* (Srimal et al., 1985). This is despite the disparity of their morphology and ecological environment. *C. rotundicauda* is approximately less than half the size of the other three species, and it is the only one that lives in brackish waters whereas the rest inhabit the open seas. In addition to other recent

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

findings on the molecular structure and enzymatic components of the coagulation enzymes, our present study on Factor C favors the widespread conservation of the system across the different species of horseshoe crabs.

Expression potentials of recombinant Factor C

In view of the harsh estuarine-brackish habitat of the *C. rotundicauda,* where extremely high populations of Gram-negative bacteria up to 10^{12} CFU/g of sediment is expected to challenge its survival, this species of horseshoe crab has a potent antibacterial defense mechanism (Yeo et al., 1993; Ding et al., 1993b; Ho et al., 1994). Indeed its amoebocyte lysate has been found to be more sensitive to endotoxin than the *Limulus* or *Tachypleus* amoebocyte lysates (Kim et al., 1988). Being the first enzyme in the coagulation cascade reaction that is activated by endotoxin, recombinant Factor C from *C. rotundicauda* would be of great significance to the pharmaceutical industry, in which pyrogen-detection constitutes an inherent part of quality assurance. The commercial implications of Factor C could be realized by its synthesis from expression vectors to produce recombinant Factor C to be used to formulate a rapid and sensitive detection kit for Gram-negative bacterial endotoxin using synthetic substrates. Subcloning the CrFC cDNAs in expression vectors and in vitro expression studies with transcription-coupled-translation (TnT) reactions in which the CrFC cDNA inserts were driven by T7 promoters have revealed higher expression of Factor C from CrFC21 constructs (our unpublished data, patent pending). In comparison with native CrFC26, the higher expression capability of CrFC21 is probably due to the absence of multiple false-start ATGs in the 5′ region, the more easily accessible AUG, and the presence of a consensus Kozak sequence in the CrFC21.

Experimental Procedures

Procurement of C. rotundicauda *amoebocytes*

The horseshoe crabs were partially bled by cardiac puncture and subsequently returned alive to their habitat. Blood was collected into equal volumes of pyrogen-free NaCl (3%) containing 0.125% *N*-ethylmaleimide (Calbiochem) as anticoagulant. Amoebocytes were pelleted by centrifugation at 150 × *g* for 30 minutes at room temperature. The cells were washed with 3% NaCl and quick-frozen in liquid nitrogen before being used for RNA preparation. All procedures were performed under pyrogen-free conditions at 0°–4°C unless otherwise stated.

Poly (A)$^-$ RNA preparation

Total RNA was purified according to Chirgwin et al. (1979). Amoebocytes of 2.5 g wet weight were resuspended in 18 ml of 4 M guanidinium isothiocynate (Fluka) containing 20 mM sodium acetate, pH 5.2, 0.1 mM DTT, and 0.5% sarkosyl. Cell lysis was enhanced by passing the suspension several times through an 18-gauge needle and four to five times through a 23-gauge needle until a homogenous lysate was obtained. The lysate was then layered onto a cushion of 4 ml of 5.7 M CsCl solution in a Beckman quickseal tube. Ultracentrifugation was effected at 400,000 × *g* for 16 hours at 18°C. The total cellular RNA pellet was air-dried and resuspended in 1 ml of TES (10 mM Tris-HCl, pH 7.4, containing 5 mM EDTA and 1% SDS) and extracted twice with chloroform:butanol (4:1). The RNA was precipitated with ethanol, resuspended in DEPC-treated water, digested with RNase-free DNase I (BRL), and extracted with phenol-chloroform. Poly(A)$^+$ RNA was purified by chromatography through Oligo-dT Tris Acryl (IBF, France).

cDNA synthesis and cloning

The mRNAs purified from amoebocytes were used to synthesize cDNAs following a modification of the method of Gubler and Hoffman (1983). *Not*I primer adaptors were used to prime the first-strand cDNA synthesis to ensure orientation-specific ligation to vector, λgt22 (Promega, USA). A replacement reaction was employed for the second-strand synthesis while *Eco*RI adaptors were used to produce sticky-ended double-stranded cDNAs. After digestion with *Not*I, the double-stranded cDNAs flanked by *Eco*RI–*Not*I restriction sites were ligated to *Eco*RI–*Not*I-digested λgt22 vector arms. The recombinant DNA was packaged and transduced into *Escherichia coli,* LE392, and Y1090. The cloning strategy is outlined in Figure 9.

Screening of λgt22 cDNA library

The λgt22 cDNA library was screened using, as probe, *T. tridentatus* Factor C cDNA (TtFC53), which is a partial fragment of the Factor C cDNA (Muta et al., 1991; a kind gift from Prof. S. Iwanaga). As a rare copy gene, *C. rotundicauda* Factor C cDNA represented only 0.03% of the amoebocyte cDNA library. A total of 40 presumptive clones were isolated, of which 16 were further studied. The *C. rotundicauda* λ Factor C recombinants (CrFC) were

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

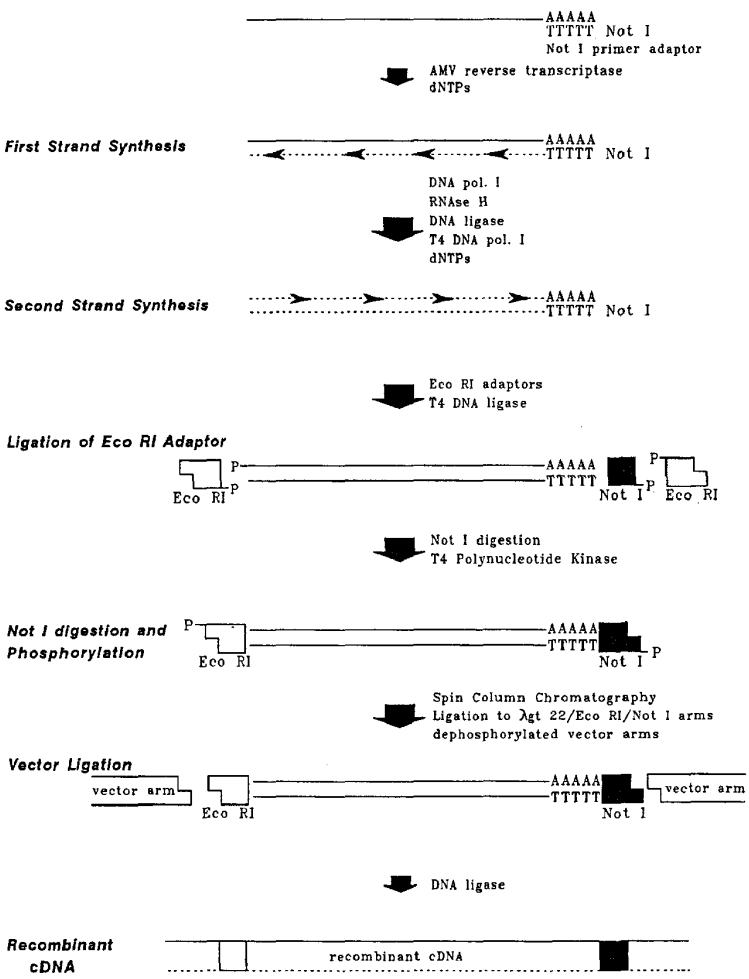


Figure 9. The overall strategy utilized for the synthesis and cloning of the amoebocyte cDNA. *Not*I primer adaptors were used to prime the first-strand synthesis and allow orientation-specific or directional ligation into the vector, bacteriophage λgt22. A replacement reaction was employed for the second-strand synthesis reaction using RNase H, DNA polymerase I, and DNA ligase while *Eco*RI adaptors were used to produce a sticky-ended double-stranded cDNA. Afer digestion with *Not*I, the unique restriction ends were phosphorylated with T_4 kinase to allow directional ligation to the dephosphorylated *Eco*RI–*Not*I-digested vector arms. After packaging in vitro, the phage particles were transduced into bacterial host cells (*E. coli,* LE392, and Y1090) for subsequent propagation or cloning.

purified, and the size of their inserts were determined after digestion with *Eco*RI and *Not*I restriction enzymes.

Northern and Southern blot analysis

RNA samples were denatured with glyoxal and Me_2SO (McMaster and Carmichael, 1977), separated on 1% agarose gel, and transferred to nylon membrane (PALL Biodyne A, 0.22 μm) (Reed and Mann, 1985). Likewise, DNA samples were fractionated on 1% agarose gels, depurinated, and transferred as above under alkaline conditions. ^{32}P-labeled DNA probes (either heterologous TtFC53 or, later, homologous CrFC cDNA) with specific activities of 10^9 cpm/μg of DNA were routinely used for hybridization under standard conditions (Maniatis et al., 1982). With the heterologous TtFC53 probe, hybridization was carried out at 42°C in the presence of 50% formamide. Posthybridization washing was effected with 1× SSC/0.1% SDS at room temperature followed by 0.1× SSC/0.1% SDS at 42°C until low background radioactivity was attained. With homologous CrFC cDNA probes, higher-stringency conditions of hybridization and washings were employed at 65°C.

Subcloning of CrFC cDNA fragments in pGEM 11Zf(+)

To facilitate further manipulations of the CrFC cDNA, the *Eco*RI-*Eco*RI (EE) and *Eco*RI-*Not*I (EN) fragments were subcloned into plasmid pGEM 11Zf(+) and transformed into *E. coli* JM109. The orientation of the EE fragments subcloned into the *Eco*RI site of pGEM 11Zf(+) was determined before restriction mapping using the unique *Xho*I site at the 3′ end of EE fragment and pGEM 11Zf(+). The

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

subclones containing EE and EN fragments are henceforth named pEE and pEN, respectively. The full-length CrFC21 and CrFC26 were also inserted into pGEM 11Zf(+).

DNA sequencing of CrFC21 and CrFC26

The DNA sequences of CrFC21 and CrFC26 were determined on both strands by the dideoxy method of Sanger et al. (1977). Figure 2 shows the sequencing strategies for CrFC26 and CrFC26. To countercheck the sequences obtained, the 5′ ends of some other CrFC subclones; CrFC1, CrFC16, CrFC35, and CrFC69 (see Figure 2) were also sequenced using T7 promoter as the primer binding site.

Acknowledgments

This work was supported by grant RP: 88/357 (B.H. and J.L.D.). M.A.A.N. III was in receipt of an National University of Singapore Research Scholarship. We thank Mr. C. Chai for technical assistance and Mr. H.K. Yip for photography.

References

Chirgwin, J.M., Przbyla, A.E., McDonald, R.J., and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294–5299.

Ding, J.L., Navas, M.A.A., and Ho, B. (1993a). Two forms of Factor C from the amoebocytes of *Carcinoscorpius rotundicauda:* purification and characterisation. *Biochim Biophys Acta* 1202:149–156.

Ding, J.L., Sababathy, T.K., and Ho, B. (1993b). Morphological changes of *Carcinoscorpius rotundicauda* amoebocyte and *Escherichia coli* during their interaction. *Cytobios* 75:21–32.

Furie, B., and Furie, B.C. (1988). The molecular basis of blood coagulation. *Cell* 53:505–518.

Gubler, U., and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* 25:263–269.

Hartley, B.S. (1970). Homologies in serine proteases. *Philos Trans R Soc Lond Biol* 257:77–87.

Ho, B. (1983). An improved *Limulus* gelation assay. *Microbios Lett* 24:81–84.

Ho, B., Kim, J.C., and Ding, J.L. (1993). Electrophoretic analysis of endotoxoin-activated gelation reaction of *Carcinoscorpius rotundicauda* amoebocyte lysate. *Biochem Mol Biol Int* 29:687–694.

Ho, B., Yeo, D.S.A., and Ding, J.L. (1994). A tetrodotoxin neutralizing system in the haemolymph of the horseshoe crab, *Carcinoscorpius rotundicauda. Toxicon* 32:755–762.

Ingolia, T.D., and Craig, E.A. (1981). Primary sequence of the 5′ flanking regions of the *Drosphila* heat shock genes in chromosome subdivision 67B. *Nucleic Acids Res* 9:1627–1642.

Iwanaga, S., Morita, T., Miyata, T., and Nakamura, T. (1985). Haemolymph coagulation system in *Limulus.* In : *Microbiology,* Leive, L., Bonventre, P.F., Morello, J.A., Schlesinger, S., Silver, S.D., and Wu, H.C. (eds.). Washington, D.C.: American Society of Microbiology, pp. 29–32.

Kim, J.C., Ding, J.L., and Ho, B. (1988). Endotoxin activation of clotting proteins from *Carcinoscorpius rotundicauda* ameobocyte lysate. In: *Advances in Biochemistry & Biotechnology in Asia and Oceania,* Sipat, A., Ampon, K., Perumal, R., Aziz, S., and Thambyrajah, V. (eds.). Malaysian Biochemical Society (FAOB), Kuala Lumpur, Malaysia: pp. F1.

Kozak, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiol Rev* 47:1–45.

Kozak, M. (1986). Bifunctional messenger RNAs in eucaryotes. *Cell* 47:481–483.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying hydropathic character of a protein. *J Mol Biol* 157:105–132.

Lehtovaara, P., Soderlund, H., Keranen, S., Petterson, R., and Kaarianen, LL. (1982). Extreme ends of the genome are conserved and rearranged in the defective interfering RNAs of Semliki Forest virus. *J Mol Biol* 156:731–748.

Leytus, S.P., Foster, D.C., Kurachi, K., and Davie, E.W. (1986). Gene for human Factor X: a blood coagulation factor whose gene organization is essentially identical with that of Factor IX and protein C. *Biochemistry* 25:5098–5102.

Li, X., Trinh, K.Y., and Hew, C.S. (1991). Expression and characterization of an active and thermally more stable recombinant antifreeze polypeptide from ocean pout, *Macrozoarces americanus,* in *Escherichia coli:* improved expression by the modification of the secondary structure of the mRNA. *Protein Eng* 4:995–1002.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor, Laboratory Press.

McMaster, G.K., and Carmichael, G.G. (1977). Analysis of single and double-stranded nucleic acids on polyacrylamide and agarose gel by using glyoxal and acridine orange. *Proc Natl Acad Sci USA* 74:4835–4838.

Muta, T., Miyata, T., Misumi, Y., Tokunaga, F., Nakamura, T., Toh, Y., Ikehara, Y., and Iwanaga, S. (1991). *Limulus* Factor C: an endotoxin sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like and lectin-like domains. *J Biol Chem* 266:6554–6561.

Nakamura, T., Morita, T., and Iwanaga, S. (1985). Intracellular proclotting enzyme in *Limulus (Tachypleus tridentatus)* hemocytes: its purification and properties. *J Biochem* 97:1561–1574.

Nakamura, T., Morita, T., and Iwanaga, S. (1986). Lipopolysaccharide sensitive serine-protease zymogen (Factor C) found in *Limulus* hemocytes: isolation and characterization. *Eur J Biochem* 154:511–521.

Nakamura, T., Tokunaga, F., Morita, T., Iwanaga, S., Kusumoto, S., Shiba, T., Kobayashi, T., and Inoue, K. (1988). Intracellular serine-protease zymogen, Factor C, from horseshoe crab hemocytes: its activation by synthetic lipid A analogues and acidic phospholipids. *Eur J Biochem* 176:89–94.

Navas, M.A.A., Ding, J.L., and Ho, B. (1990). Inactivation of

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Factor C by dimethyl sulfoxide inhibits coagulation of the *Carcinoscorpius* amoebocyte lysate. *Biochem Int* 21:805–813.

Park, C.H., and Tulinsky, A. (1986). Three-dimensional structure of the kringle sequence: structure of prothrombin-fragment I. *Biochemistry* 25:3977–3982.

Reed, R.C., and Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res* 13:7207–7221.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.

Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R, Fong, N., Selby, M., Rutter, W.J., and Bell, G.I. (1983). Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins. *Science* 221:236–240.

Solum, N.O. (1973). The coagulogen of *Limulus polyphemus* hemocytes, a comparison of the clotted and the non-clotted forms of the molecule. *Thromb Res* 2:55–70.

Srimal, S., Miyata, T., Kawabata, S., Miyata, T., and Iwanaga, S. (1985). The complete amino acid sequence of coagulogen isolated from Southeast Asian horseshoe crab, *Carcinoscorpius rotundicauda. J Biochem* 98:305–318.

Watt, R., Stanton, L.W., Marcu, K.B., Gallo, R.C., Croce, C.M., and Rovera, G. (1983). Nucleotide sequence of cloned cDNA of human c-myc oncogene. *Nature* 303:725–728.

Yeo, D.S.A., Ding, J.L., and Ho, B. (1993). An antimicrobial factor from the plasma of the horseshoe crab, *Carcinoscorpius rotundicauda. Microbios* 73:45–58.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)