© 1990 Oxford University Press

DNA recombination during PCR

Andreas Meyerhans, Jean-Pierre Vartanian and Simon Wain-Hobson*
Laboratoire de Biologie et Immunologie Moléculaires des Rétrovirus, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France

Received March 1, 1990; Accepted March 9, 1990

ABSTRACT

PCR co-amplification of two distinct HIV1 *tat* gene sequences lead to the formation of recombinant DNA molecules. The frequency of such recombinants, up to 5.4% of all amplified molecules, could be decreased 2.7 fold by a 6 fold increase in Taq DNA polymerase elongation time. Crossover sites mapped essentially to three discrete regions suggesting specific Taq DNA polymerase pause or termination sites. PCR mediated recombination may be a problem when studying heterogeneous genetic material such as RNA viruses, multigene families, or repetitive sequences. This phenomenon can be exploited to create chimeric molecules from related sequences.

INTRODUCTION

It is assumed that the products after polymerase chain reaction (PCR; refs. 1–3) are homogeneous. This is usually so. However when target sequences, such as multigene families, alleles or RNA viruses are amplified, the PCR products will be heterogeneous. In such instances recombinant or 'shuffled' molecules, i.e. molecules composed of parts of two different sequences artificially combined during PCR may be produced (4). However the frequency or the factors influencing the generation of these PCR artifacts are not known.

PCR mediated recombinants presumably arise due to the presence of a proportion of incompletely extended primers annealing to a heterologous target sequence. Incomplete DNA strand synthesis has recently been described (5) and is enshrined in standard PCR protocols in which the last step is a 10 minute polymerization reaction. Despite a certain degree of nucleotide mismatching present in such partially double stranded molecules they will probably be completed in the subsequent elongation step. Indeed it has been shown that even a mismatch of the 3′ base in the primer may result in strand synthesis (6).

Our interest in PCR mediated recombination stems from the fact that PCR has been extensively used to characterize human immunodeficiency virus type 1 (HIV-1) proviral populations which are very heterogenous (7,8). During replication HIV, like other RNA viruses, exhibits a high nucleotide misincorporation rate. Thus HIV-infected persons always harbour a cluster of related viral genomes. Consequently single viral sequences are not sufficient to adequately describe any viral strain and a population based approach should be used. By cloning PCR amplified DNA and sequencing individual clones viral populations were established (7).

For such a strategy to be valid certain requirements have to be fulfilled: 1) the amplification primers should pick up all of the sequences present; 2) the cloning and selection of clones for sequencing should be unbiased; 3) the intrinsic Taq DNA polymerase misincorporation rate during primer elongation reactions should be significantly lower than the variation within the gene or locus to be studied and 4) the frequency of PCR mediated recombination should be low.

The first three points have already been addressed specfically for HIV1 (6–8). In this paper the problem of PCR mediated recombination between related sequences is addressed. For a 242 base pair segment of the HIV1 *tat* gene the frequency of recombinant or shuffled clones is no greater than 5%.

MATERIALS AND METHODS

Plasmids

Plasmids pSV2tatBru and pSV2tatMal contain the first coding exon of the *tat* genes derived from molecular clones of HIV-1 strain Bru and HIV-1 strain Mal respectively (9,10). The latter plasmid was constructed in a similar fashion as previously described for pSV2tatBru (7). Since the same PCR primers, namely T1 and T2, were used for their construction both HIV regions are flanked by the same sequences.

Oligodeoxynucleotides

All oligodeoxynucleotides were synthesized on an Applied Biosystems DNA synthesizer (model 380) using phosphoramidate chemistry. After detritylation and alkaline deprotection, crude material was used for all applications without any further purification. The T1 and T2 primers have already been described (7). The sequences of the hybridization primers B1 and B2 (HIV1 Bru specific) and M1 and M2 (HIV1 Mal specific) are shown in Figure 1.

PCR and M13 cloning

Ten nanograms (approximately 2×10^9 molecules) of plasmid pSV2tatBru and pSV2tatMal DNA were mixed and subjected to 35 cycles of amplification with primers T1 and T2 in a Perkin Elmer-Cetus thermal cycler. Reactions contained 2.5mM

* To whom correspondence should be addressed

```
        probe B1        region 1        region 2        region 3
BRU   GTTACTCAAC AGAGGAGAGC AAGAAATGGA GCCAGTAGAT CCTAGACTAG AGCCCTGGAA GCATCCAGGA  70
MAL   A......G.. ....AG.... .......... T......... ....ACT... .......... C........G
        probe M1

      region 4                                                  region 5
      AGTCAGCCTA AAACTGCTTG TACCACTTGC TATTGTAAAA AGTGTTGCTT TCATTGCCAA GTTTGTTTCA 140
      .......... GG..GC.... ..AT.AG..T .......... ....C....A .......... A.G..C....

                                        region 6
      CAACAAAAGC CTTAGGCATC TCCTATGGCA GGAAGAAGCG GAGACAGCGA CGAAGACCTC CTCAAGGCAG 210
      T...G....G .......... .......... .......... .......... .......... ....G....A

                                                                  probe B2
      TCAGACTCAT CAAGTTTCTC TATCAAAGCA GTAAGTAGTA CATGTAATGC AACCTATACA AATAG      275
      ....G..... ....A.C... ..C..G.... .......... T.......A. .....T..GT G...T
                                                                  probe M2
```

Figure 1. Nucleotide sequences of the HIV1 *tat* gene for strains Bru and Mal. Only the nucleotide differences are given for the Mal sequence, identities being scored by a dot. The strain specific hybridization primers B1 and B2 (strain Bru) and M1 and M2 (strain Mal) are indicated. The amplification primers T1 and T2 flank the 5′ and 3′ ends respectively of the sequences. The six regions within which PCR mediated recombination occurred are also indicated. The *tat* initiator methionine and stop codons are located at positions 26–28 and 242–244 respectively.

$MgCl_2$, 50mM KCl, 10mM Tris-HCl pH 8.3, 0.2mM dNTPs each, 50pmol of each primer, 2.5U Taq DNA polymerase (Cetus, approximately 64×10^9 molecules initially) in a final volume of 100μl and were overlaid with 50μl of heavy white mineral oil (Sigma). Thermal cycling conditions were 1 min. at 95°C, 1 min. at 55°C and 30 sec. or 3 min. at 72°C. Amplification was completed by a final incubation at 72°C for 10 min. PCR products were purified from 5% polyacrylamide gels, cleaved with HindIII and KpnI (which cut only in the T1 and T2 amplification primer sequences) and ligated to similarily digested and dephosphorylated M13mp18 RF DNA. After transformation of *E.coli* strain TG1 white plaques from X-Gal/IPTG plates were screened for Bru/Mal or Mal/Bru recombinants.

Screening for Bru-Mal recombinants

About 500 white M13 plaques from each of the 30 second and 3 minute PCR reactions were repicked in quadruplicate on LB plates covered with TG1. After transfer to nitrocellulose, filters were prehybridized for 1 h at 37°C in 5×SSC and 1× Denhardt's solution. Filters were then transferred to a fresh solution (as above) containing a further 2mM EDTA and 5′-^{32}P end-labelled and heat-denatured primer (10^6 cpm/ml hybridization solution). Each filter was hybridized with only one of the four probes for 1h at 37°C. The specificity of these HIV1 Bru and Mal probes was first tested on M13 clones carrying either the corresponding Bru or Mal target sequences. The hybridization and washing conditions were such that the probes were completely specific. Thus filters were washed 3 times in 2×SCC, 0.1% SDS at 37°C and exposed to X-ray film overnight. M13 plaques giving positive hybridization signals with B1 and M2 or M1 and B2 primers were picked and prepared for DNA sequencing.

Localisation of crossover sites

All Bru/Mal recombinant clones were sequenced with the universal M13 primer by standard dideoxy methods using (α-^{35}S) dATP (600 Ci/mmol) and resolved on buffer gradient gels as previously described (11).

RESULTS

An equimolar mixture of plasmids pSV2tatBru and pSV2tatMal (10ng each) was subjected to 35 cycles of PCR under standard conditions using 30 sec. or 3 min. elongation times. Amplified *tat* gene sequences were cloned into M13mp18 and screened for recombinants. The rationale for using a relatively high initial DNA concentration was the following: recombinant PCR products probably arise when partially extended primers anneal to heterologous target sequences and are elongated. For such an event to occur the abortive extension products must compete successfully with the normal PCR amplification primers for target sequences. The probability of this occuring will increase with each subsequent round of amplification as the ratio of normal PCR primer to target DNA progressively decreases. By using a high initial plasmid concentration these conditions would be reached relatively early on in the reaction, consequently favouring recombination. Thus recombinant frequencies derived from these experiments would represent upper estimates for such events in an otherwise typical PCR reaction.

Illumina Exhibit 1018, Page 2 of 5

Table 1. Frequencies of region specific PCR mediated recombinants. The six regions are described in Fig. 1. The number of recombinants, their relative proportion and the recombination frequency per base (i.e. the number of recombinants/region normalized to the length of that region) are given separately for the 30 second and 3 minute experiments. The combined totals are shown on the right of the table.

Region	1	2	3	4	5	6	
Region length (bases)	14	13	13	10	10	54	Total
Elongation time 30 s.							
No. Recombinants	4	3	1	1	1	17	27
% Recombinants	15	11	4	4	4	62	100
Frequency/base	0.29	0.23	0.08	0.10	0.10	0.31	-
Elongation time 3 m.							
No. Recombinants	2	1	0	0	0	7	10
% Recombinants	20	10	0	0	0	70	100
Frequency/base	0.14	0.08	0	0	0	0.13	-

The nucleotide sequence of the *tat* genes from HIV1 isolates Bru and Mal are shown in Figure 1. In addition the strain specific hybridization probes B1 and B2 (HIV1 Bru) and M1 and M2 (HIV1 Mal) used in the study are also shown. The PCR amplification primers T1 and T2 used to amplify the Bru and Mal *tat* sequences flank the 5′ and 3′ ends of the sequences respectively.

A total of 37 plaques were positive on hybridization with B1 and M2 or M1 and B2 and therefore considered recombinant molecules. Twenty-seven (5.4% of all plaques picked) were identified when the elongation time was 30 seconds and only ten (2.1% of all plaques picked) when the elongation time was 3 minutes. The comparable frequencies of B1/M2 and M1/B2 recombinants in both experiments suggested that recombinant formation was unbiased (data not shown).

In order to localize the crossover points all 37 recombinants were sequenced. Given that the HIV1 Bru and Mal *tat* genes differed by 32 bases (13%) between the hybridization probes, these sites could be mapped to within 10–54 bases depending on the segment (Fig. 1). The number and proportion of recombinants as identified by their crossover regions are given in Table 1. While there were fewer recombinants detected using a longer elongation time their relative proportion for a given region remained essentially the same. All 37 recombinants encoded a single recombination site. However the length of the six crossover regions varied considerably (Fig. 1). Thus the recombinant frequency per base pair (i.e. the number of recombinants/ region normalized to the length of the region) is also given in Table 1. This value is comparable for regions 1, 2 and 6, and significantly less for regions 3, 4 and 5. This suggested that there were preferred regions in the *tat* gene sequence for PCR mediated recombination.

The HIV-1Bru and Mal sequences between the T1 and T2 amplification primers were examined for secondary structure using the programmes of Zuker (12). No obvious or stable secondary structures coincided with the recombination regions.

DISCUSSION

That recombinant structures were identified meant that incompletely elongated DNA strands were produced in the PCR reactions even after an elongation time of 3 minutes. Their formation could be due to one of three possibilities. Firstly, Taq DNA polymerase could be simply pausing on the template. This event must be of the order of a minute or so such that continued strand synthesis may be interrupted by thermal cycling. Secondly, it could be due to premature termination. This would be consistent with the fact that Taq DNA polymerase is only a moderately processive enzyme.

A third possibility is that there may be preferred sites of

```
         10         20         30         40         50         60         70
MEPVDPRLEP WKHPGSQPKT ACTTCYCKKC CFHCQVCFTT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQ     BRU
.......... .......... .......... .......... .......... .......... .N.A..DP.P E.
.......... .......... .......... .....M..I. .G........ .......... .N.A..DP.P E.
.......... ........R. P.NK...... .Y...M..I. .G........ .......... .N.A..DP.P E.  5'BRU/3'MAL
.......... .N......R. P.NK...... .Y...M..I. .G........ .......... .N.A..DP.P K.
......N... .N......R. P.NK...... .Y...M..I. .G........ .......... .N.A..DP.P K.
.D....N... .N......R. P.NK...... .Y...M..I. .G........ .......... .N.A..DP.P E.     MAL
.D....N... .N......R. P.NK...... .Y...M..I. .G........ .......... .......... ..
.D....N... .N......R. P.NK...... .Y........ .......... .......... .......... ..
.D....N... .N........ .......... .......... .......... .......... .......... ..  5'MAL/3'BRU
.D....N... .......... .......... .......... .......... .......... .......... ..
.D........ .......... .......... .......... .......... .......... .......... ..
1   2    3     4                   5                   6                        region
```

Figure 2. HIV1 *tat* protein chimeras produced by PCR mediated recombination. All sequences, in the one letter code, are given relative to the HIV1 Bru sequence. Only amino acid difference have been scored. Thus the first four sequences represent 5′ Bru/3′Mal recombinants. The HIV1 Mal sequence and four 5′Mal/3′Bru recombinants are shown below. The one letter code is: A Ala; C Cys; D Asp; E Glu; F Phe; G Gly; H His; I Ile; K Lys; L Leu; M Met; N Asn; P Pro; Q Gln; R Arg; S Ser; T Thr; V Val; W Trp; Y Tyr.

nucleation between complementary DNA strands. During the renaturation and elongation steps this may give rise to partially double stranded structures leading to blocking of the Taq DNA polymerase. If this was really operative one would have expected more recombinants in the case of the 3 minute elongation reaction as compared to the 30 second reaction. That the converse was found would appear to argue against this. Indeed the observation is consistent with either of the first two possibilities.

Despite the lack of any obvious secondary structure it was possible that regions 1, 2 and 6 in the *tat* sequence in question were particularly vulnerable to PCR mediated recombination and therefore not representative. In order to see whether recombinants could be identified at comparable frequencies within another sequence, a 111 base pair fragment in the first hypervariable region of the HIV1 *env* gene was studied. Two distinct *env* sequences (clones B40.01 and B40.02, ref. 8) were chosen. Using identical PCR conditions as in above experiments (elongation time 30 sec.) the overall recombinant frequency turned out to be a factor of six less when normalized to the comparable *tat* experiment. This means that recombinant frequencies are of the order of 1−5% suggesting that PCR mediated recombination is a general phenomenon.

On a practical level the frequency of HIV1 *tat* recombinants does not significantly affect the strategy adopted in describing HIV1 proviral populations (7). In this work 20 clones of each sample were sequenced yielding a resolution of 5%. These studies incorporated an elongation time of 1 minute which would place the resolution of the study just above the threshold of recombinant formation. Yet given the highly unfavourable conditions used here (i.e. high input DNA concentration) the recombinant frequencies described in this paper must represent upper limits.

It is to be expected that recombinant formation will be a function of the particular sequence (specific pause or termination sites), the length of target DNA (by virtue of encoding more sites), the elongation time (shown here) and initial DNA and Taq DNA polymerase concentration (presumably). Indeed conditions could be 'optimized' for the recovery of recombinants. This could be achieved by reducing the primer concentration, conducting only two cycle PCR with a short annealing/elongation time and ensuring that the molar ratio of input DNA to enzyme was not too low. In so doing a wide range of recombinant molecules could be generated. Even in this 'unoptimized' study a total of 10 HIV1 *tat* gene recombinants were produced judging from the amino acid sequences (Fig. 2). Thus PCR mediated recombination can even be exploited to produce a series of chimeric molecules in a single experiment.

Finally this study points out the vagaries of carrying out PCR for an excessive number of cycles. Such recombinant sequences are not expected to be problematic if lower concentrations of template and fewer cycles are used. Carrying out PCR of heterogeneous target DNA with short elongation times (i.e. two cycle PCR) is strongly argued against.

ACKNOWLEDGEMENTS

We would like to thank Drs. Stewart Cole and John Sninsky for critical suggestions, Rémi Cheynier for art work and Annick Galliot for typing the manuscript. A.M. is a post-doctoral research fellow of the Centre International des Etudiantes et Stagiares. This work was supported, in part, by grants from Institut Pasteur and l'Agence Nationale pour la Recherche sur le SIDA.

REFERENCES

1. Saiki, R., Sharf, S., Faloona, F., Mullis, K.B., Horn, G., Erlich, H.A. and Arnheim, N. (1985) Science **230**, 1350−1354.
2. Mullis, K.B. and Faloona, F. (1987) Methods Enzymol. **155**, 335−350.
3. Erlich, H.A., Gibbs, R. and Kazazian, H.H. (Eds.) (1989) Polymerase chain reaction. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
4. Saiki, R., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science, **239**, 487−491.
5. Olsen, D.B. and Eckstein, F. (1989) Nuc. Acids Res. **17**, 9613−9620.
6. Kwok, S., Kellogg, D.E., Levenson, C., Spasic, D., Goda, L. and Sninsky, J.J. Nuc. Acids Res. (in press).

7. Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Asjö, B. and Wain-Hobson, S. (1989) Cell, **58**, 901–910.
8. Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J. and Wain-Hobson, S. (1989) J. AIDS **2**, 344–352.
9. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S.T., and Alizon, M. (1986). Cell, **40**, 9–17.
10. Alizon, M., Wain-Hobson, S. Montagnier, L. and Sonigo, P. Cell, **46**, 63–74.
11. Biggen, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA, **80**, 3969–3965.
12. Zucker, M. and Stiegler, P. (1981) Nuc. Acids Res. **9**, 133–148.