FISH.
FISH & RICHARDSON

Street Address
Fish & Richardson P.C.
3200 RBC Plaza
60 South Sixth Street
Minneapolis, MN 55402

Mail Address
P.O. Box 1022
Minneapolis, MN 55440-1022

612 335 5070 main
877 769 7945 fax

August 16, 2017

Attorney Docket No.: 38767-0048002

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Presented for filing is a new Continuation patent application for Track I Prioritized Examination of:

Inventor(s): SMITA SHANKAR AND MARTIN ANDREW HOYT

Title: EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST

Enclosed are the following papers, including those required to receive a filing date under 37 CFR § 1.53(b):

	Pages
Certification and Request for Prioritized Examination (Track I)	1
Application Data Sheet	9
Specification	46
Claims	2
Abstract	1
Drawing(s)	11
Nucleotide and/or amino acid sequence listing (.txt format)	
Power of Attorney	2
Information disclosure statement,	1
PTO-1449	2
References	0

Basic Filing Fee	$280
Search Fee	$600
Examination Fee	$720
Publication fee	$0
Track I processing fee	$140
Track I prioritized examination fee	$4000

FISH.
FISH & RICHARDSON

Commissioner for Patents
August 16, 2017

Total Claims (13)	over 20	0 x $80	$0
Independent Claims (1)	over 3	0 x $420	$0
Fee for Multiple Dependent claims			$0
Application size fee for each 50 pages over 100			$0
Surcharge for filing fee, search fee, examination fee or oath or declaration			$0
Total Filing fee			**$5740**

The filing fee is being paid concurrently on the Electronic Filing System (EFS) by Deposit Account authorization. Please apply any other fees or credits to Deposit Account No. 06-1050, referencing the Attorney Docket Number shown above.

I hereby authorize the USPTO to communicate via email in response to email communications from the undersigned and practitioners under Customer No. 26191 in accordance with 37 CFR § 1.33 and 37 CFR § 1.34 concerning any subject matter of this application. I understand that a copy of these communications will be made of record in the application file. (MPEP § 502.03).

If this application is considered to be incomplete, or if a telephone conference would otherwise be helpful, please call the undersigned at (612) 335-5070.

Please send all correspondence to:

26191
Customer Number

Respectfully submitted,

/M. Angela Parsons/

M. Angela Parsons, Ph.D.
Registration No. 44,282

Enclosures
MAP/jxw

61545150.doc

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST
First Named Inventor/Applicant Name:	Smita Shankar
Filer:	M. Angela Parsons/Judith Witzel
Attorney Docket Number:	38767-0048002

Filed as Large Entity

Filing Fees for Track I Prioritized Examination - Nonprovisional Application under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
UTILITY APPLICATION FILING	1011	1	280	280
UTILITY SEARCH FEE	1111	1	600	600
UTILITY EXAMINATION FEE	1311	1	720	720
REQUEST FOR PRIORITIZED EXAMINATION	1817	1	4000	4000
Pages:				
Claims:				
Miscellaneous-Filing:				
PROCESSING FEE, EXCEPT PROV. APPLS.	1830	1	140	140

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**5740**

Electronic Acknowledgement Receipt

EFS ID:	30092486
Application Number:	15678342
International Application Number:	
Confirmation Number:	1018
Title of Invention:	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST
First Named Inventor/Applicant Name:	Smita Shankar
Customer Number:	26191
Filer:	M. Angela Parsons/Christin Schwartz
Filer Authorized By:	M. Angela Parsons
Attorney Docket Number:	38767-0048002
Receipt Date:	16-AUG-2017
Filing Date:	
Time Stamp:	12:07:16
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$5740
RAM confirmation Number	081617INTEFSW00009758061050
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	TrackOne Request	CertandReqT1.pdf	146178 63c770007e60293bd7f03e9aa2536a1a21e7d27b	no	2
Warnings:					
Information:					
2	Application Data Sheet	ADS.pdf	1823422 060d4763514d8450474e335d1fca68e85b747854	no	9
Warnings:					
Information:					
3		APPLN.pdf	286386 e6fbef98d81eceeb52f3c5dad56bc4a63f224a73	yes	49

Multipart Description/PDF files in .zip description

Document Description	Start	End
Specification	1	46
Claims	47	48
Abstract	49	49

Warnings:					
Information:					
4	Drawings-only black and white line drawings	dRAWING.pdf	1077684 eecc7d6ccef86ef381e2cce009fb377cafe3f152	no	11
Warnings:					
Information:					
5	Power of Attorney	TransPOA.pdf	2170394 01c8ddabc780676f54e97836be9023f7de24fdf7	no	3

Warnings:					
Information:					
6	Sequence Listing (Text File)	SEQ.txt	44696 [redacted]	no	-
Warnings:					
Information:					
7	Transmittal of New Application	TransLtr.pdf	115488 aa7cf6590660697fd24cc7c210e0d15127f9699f	no	2
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	38375 1e029ed4bd3f1808832bb0f673915439c3eb9e37	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	5702623		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRACK1.REQ
Document Description: TrackOne Request

PTO/AIA/424 (03-14)

CERTIFICATION AND REQUEST FOR PRIORITIZED EXAMINATION UNDER 37 CFR 1.102(e) (Page 1 of 1)

First Named Inventor:	Smita Shankar	Non provisional Application Number (if known):	
Title of Invention:	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST		

APPLICANT HEREBY CERTIFIES THE FOLLOWING AND REQUESTS PRIORITIZED EXAMINATION FOR THE ABOVE-IDENTIFIED APPLICATION.

1. The processing fee set forth in 37 CFR 1.17(i)(1), the prioritized examination fee set forth in 37 CFR 1.17(c), and if not already paid, the publication fee set forth in 37 CFR 1.18(d) have been filed with the request. The basic filing fee, search fee, and examination fee are filed with the request or have been already been paid. I understand that any required excess claims fees or application size fee must be paid for the application.

2. I understand that the application may not contain, or be amended to contain, more than four independent claims, more than thirty total claims, or any multiple dependent claims.

3. The applicable box is checked below:

I. ☒ Original Application (Track One) - Prioritized Examination under § 1.102(e)(1)

i. (a) The application is an original non provisional utility application filed under 35 U.S.C. 111(a). This certification and request is being filed with the utility application via EFS-Web.

---OR---

(b) The application is an original non provisional plant application filed under 35 U.S.C. 111(a). This certification and request is being filed with the plant application in paper.

ii. An executed inventor's oath or declaration under 37 CFR 1.63 or 37 CFR 1.64 for each inventor, **or** the application data sheet meeting the conditions specified in 37 CFR 1.53(f)(3)(i) is filed with the application.

II. ☐ Request for Continued Examination - Prioritized Examination under § 1.102(e)(2)

i. A request for continued examination has been filed with, or prior to, this form.
ii. If the application is a utility application, this certification and request is being filed via EFS-Web.
iii. The application is an original non provisional utility application filed under 35 U.S.C. 111(a), or is a national stage entry under 35 U.S.C. 371.
iv. This certification and request is being filed prior to the mailing of a first Office action responsive to the request for continued examination.
v. No prior request for continued examination has been granted prioritized examination status under 37 CFR 1.102(e)(2).

Signature /M. Angela Parsons/	Date 8/16/2017
Name (Print/Typed) M. Angela Parsons, Ph.D.	Practitioner Registration Number 44,282

Note: *This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

☐ *Total of _____ forms are submitted.

61545146.doc

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant *(i.e.,* GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2:

☐ Portions or **all** of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Smita		Shankar	

Residence Information (Select One) ● US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
Millbrae	CA	US

Mailing Address of Inventor:

Address 1	120 S. El Camino Real, Apt 115
Address 2	

City	State/Province	Postal Code	Country
Millbrae	CA	94030	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Martin	Andrew	Hoyt	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
San Francisco	CA	US

Mailing Address of Inventor:

Address 1	338 Garfield Street
Address 2	

City	State/Province	Postal Code	Country
San Francisco	CA	94132	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	26191		
Email Address	apsi@fr.com	Add Email	Remove Email

Application Information:

Title of the Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST		
Attorney Docket Number	38767-0048002	**Small Entity Status Claimed**	☐
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	11	**Suggested Figure for Publication (if any)**	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	● Customer Number	US Patent Practitioner	○ Limited Recognition (37 CFR 11.9)
Customer Number	26191		

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Continuation of	PCTUS2016031797	2016-05-11
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62313491	2016-03-25
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62236506	2015-10-02
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62222388	2015-09-23
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62220366	2015-09-18
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62203052	2015-08-10
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62185921	2015-06-29
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62183074	2015-06-22

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS2016031797	Claims benefit of provisional	62159899	2015-05-11
Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button.			Add

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

			Remove
Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)
Additional Foreign Priority Data may be generated within this form by selecting the **Add** button.			Add

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1 Remove

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

● Assignee	Legal Representative under 35 U.S.C. 117	Joint Inventor
Person to whom the inventor is obligated to assign.	Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor:

If the Applicant is an Organization check here. ☒

Organization Name	Impossible Foods Inc.		
Mailing Address Information For Applicant:			
Address 1	525 Chesapeake Drive		
Address 2			
City	Redwood City	**State/Province**	CA
Country	US	Postal Code	94063
Phone Number		Fax Number	
Email Address			

Additional Applicant Data may be generated within this form by selecting the Add button. Add

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

Remove

If the Assignee or Non-Applicant Assignee is an Organization check here. ☐

Prefix	**Given Name**	Middle Name	**Family Name**	Suffix

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1			
Address 2			
City		**State/Province**	
Country i		Postal Code	
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add

Signature:

Remove

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/M. Angela Parsons/			Date (YYYY-MM-DD)	2017-08-16
First Name	M. Angela	Last Name	Parsons	Registration Number	44282

Additional Signature may be generated within this form by selecting the Add button. Add

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	38767-0048002
	Application Number	
Title of Invention	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST	

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1 The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3 A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent CooperationTreaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

F&R Ref No.: 38767-0048002

EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST

CLAIM OF PRIORITY

This application is a continuation and claims the benefit of priority to PCT/US2016/031797 filed May 11, 2016, which claims the benefit of priority to U.S. Provisional application 62/313,491 filed march 25, 2016, U.S. Provisional Application 62,236,506 filed October 2, 2015, U.S. Provisional Application 62/222,388 filed September 23, 2015, U.S. Provisional Application 62/220,366 filed September 18, 2015, U.S. Provisional Application No. 62/203,052 filed August 10, 2015, U.S. Provisional Application No. 62/185,921 filed June 29, 2015, U.S. Provisional Application No. 62/183,074 filed June 22, 2015, and U.S. Provisional Application No. 62/159,899 filed May 11, 2015, the entire contents of which are hereby incorporated by reference.

TECHNICAL FIELD

This disclosure generally relates to DNA constructs and methods of using such DNA constructs to genetically engineer methylotrophic yeast.

BACKGROUND

Methylotrophic yeast such as *Pichia pastoris* are commonly used for expression of recombinant proteins. Constructs that can be used to efficiently express one or more polypeptides in a methylotrophic yeast are provided herein.

SUMMARY

This disclosure describes the use of *P. pastoris* strains that overexpress the transcriptional activator, Mxr1, from the AOX1 promoter to increase expression of transgenes that also are expressed from the AOX1 promoter, which significantly improves the recombinant production of one or more proteins. In addition, expression of Mxr1 from the AOX1 promoter creates a positive feedback loop that allows for expression of other transgenes from the AOX1 promoter in the

absence of methanol, the normally obligate inducer, when repressing carbon sources are depleted. Expression of Mxr1 results in a significant increase in the amount of protein produced.

In one aspect, a methylotrophic yeast cell is provided that includes a recombinant nucleic acid molecule. The recombinant nucleic acid molecule typically includes an exogenous nucleic acid encoding a transcriptional activator operably linked to at least one methanol-inducible promoter element. Representative methylotrophic yeast can be of the genus *Candida*, *Hansenula*, *Pichia* or *Toruplosis*. A representative methylotrophic yeast is *Pichia pastoris*.

In some embodiments, the recombinant nucleic acid molecule is stably integrated into the genome of the methylotrophic yeast cell. In some embodiments, the recombinant nucleic acid molecule is extrachromosomally expressed from a replication-competent plasmid.

In some embodiments, the exogenous nucleic acid encoding a transcriptional activator comprises a Mxr1 sequence from *Pichia pastoris*, a Adr1 sequence from *Hansenula polymorpha*, a Trm1 sequence from *Candida boidinii*, and a Trm2 sequence from *Candida boidinii*. A representative nucleic acid encoding a transcriptional activator is shown in DQ395124. A representative transcriptional activator has an amino acid sequence shown in ABD57365.

In some embodiments, the at least one methanol-inducible promoter element is an alcohol oxidase 1 (AOX1) promoter element from *Pichia pastoris*, an AOD1 promoter element from *Candida boidinii*, a MOX promoter element from *Hansenula polymorpha*, a MOD1 promoter element from *Pichia methanolica*, a DHAS promoter element from *Pichia pastoris*, a FLD1 promoter element from *Pichia pastoris*, or a PEX8 promoter element from *Pichia pastoris*.

In some embodiments, the methylotrophic yeast cell further includes a nucleic acid molecule that includes at least one heterologous nucleic acid encoding a polypeptide operably linked to at least one methanol-inducible promoter element. In some embodiments, the at least one heterologous nucleic acid encodes one or more polypeptides involved in the biosynthesis of an iron co-factor such as heme (e.g., ALA synthase, ALA dehydratase, porphogilinogen deaminase, UPG III synthase, UPG III decarboxylase, CPG oxidase, PPG oxidase, and/or ferrochelatase). In some embodiments, one or more of the polypeptides involved in the biosynthesis of the iron co-factor are linked to at least one methanol-inducible promoter element.

In another aspect, a method for expressing a heterologous polypeptide in a cell is provided. Such a method typically includes providing a methylotrophic yeast cell as described herein; introducing a recombinant nucleic acid molecule into methylotrophic yeast cell, the

recombinant nucleic acid molecule comprising at least one heterologous nucleic acid encoding a polypeptide operably linked to at least one *Pichia pastoris* alcohol oxidase 1 (AOX1) promoter element; and culturing the cell under conditions suitable for expression of the recombinant nucleic acid molecule, thereby expressing the heterologous polypeptide.

In some embodiments, the conditions under which the cells are cultured includes the addition of iron or a pharmaceutically or metabolically acceptable salt thereof. In some embodiments, the introducing step includes a technique such as transduction, electroporation, biolistic particle delivery, or chemical transformation. In some embodiments, the culturing step includes culturing the cell in the present of methanol.

In another aspect, a recombinant organism is provided that includes a transcriptional activator operably linked to the promoter it activates. In some embodiments, a recombinant organism is provided that expresses a polypeptide operably linked to the promoter. In yet another aspect, a method of expressing a polypeptide from an inducible promoter without addition of an inducer is provided as described herein.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the methods and compositions of matter belong. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the methods and compositions of matter, suitable methods and materials are described below. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

DESCRIPTION OF DRAWINGS

Figure 1 is a schematic depicting the steps involved in the heme biosynthesis pathway.

Figure 2 are schematics of plasmids used in the construction of production strain MXY0183.

Figure 3 is a schematic showing the generation of the production strains, MXY0183 and MXY0207, from the parent strain, Bg11.

Figure 4 are schematics showing plasmids pGAB and pMx354.

Figure 5 is a schematic showing the generation of the antibiotic selection free production strains, MXY0291 and MXY0338, from the parent strain, Bg11.

Figure 6 is a schematic of the linear pieces of DNA containing Mxr1 and LegH var 3 that were introduced by co-transformation to make the production strain MXY0291.

Figure 7 is a schematic showing the linear construct expressing LegH under control of native *Pichia* non-pAOX1 constitutive promoters.

Figure 8 is a photograph showing the phenotypic changes associated with strain, MXY0183. Shake flasks at the start of induction (0 hr) and 72 hr post-induction are shown. 1, MXY0051; 2, MXY0118; 3, MXY0183.

Figure 9 shows the production of LegH from the modified *P. pastoris* strains. Panel A is a SDS gel showing lysates from *P. pastoris* strains grown in shake flasks: 51, MXY0051; 118, MXY0118; 183, MXY0183. Panel B is a table comparing LegH production from strains MXY0118, MXY0183, and MXY0207.

Figure 10 shows data from experiments with strain MXY0206. Panel A is a photograph of shake flask cultures of strains MXY0183 (left) and MXY0206 (right) after 48 hr of growth in repressing carbon source. Panel B is a photograph of cell pellets from shake flask cultures of strains MXY0183 (left) and MXY0206 (right) after 48 hr of growth in BMY media. Panel C is a graph showing the relative yield of heme-loaded LegH (in the absence of any induction agent).

Figure 11 is a summary table showing relative yields of strains described herein when grown in the presence of methanol with glycerol or methanol with glucose in 2 L fermenter tanks.

DETAILED DESCRIPTION

Nucleic acid constructs are provided herein that allow for genetically engineering a cell to increase the recombinant expression of a polypeptide. In some embodiments, nucleic acid constructs are provided herein that allow for genetically engineering a cell to increase the recombinant expression of a polypeptide from an inducible promoter in the absence of the inducing molecule. Without being bound by any particular mechanism, the methods described herein create a positive feedback loop where the low level native expression of a transcriptional activator induces a promoter that is operably linked to a transcriptional activator. This leads to an increased expression of the transcriptional activator as well as one or more target polypeptides that are operably linked to the same inducible promoter.

Nucleic acid constructs are provided herein that allow for genetically engineering a methylotrophic yeast cell. While the methods are exemplified herein using a *Pichia* species (i.e., *P. pastoris*), other species of the *Pichia* genus can be used or species from any of the *Candida*, *Hansenula*, *Pichia* and *Torulopsis* genera.

Genetically engineering a methylotrophic yeast cell typically includes introducing a recombinant nucleic acid molecule into the cell. As described herein, a recombinant nucleic acid molecule typically includes an exogenous nucleic acid that encodes a transcriptional activator operably linked to at least one inducible promoter element.

Recombinant nucleic acid molecules used in the methods described herein are typically DNA, but RNA molecules can be used under the appropriate circumstances. As used herein, "exogenous" refers to any nucleic acid sequence that is introduced into the genome of a cell from an external source, where the external source can be the same or a different organism or a nucleic acid generated synthetically. For example, an exogenous nucleic acid can be a nucleic acid from one microorganism (e.g., one genus or species of methylotrophic yeast) that is introduced into a different genus or species of methylotrophic yeast, however, an exogenous nucleic acid also can be a nucleic acid from a methylotrophic yeast that is introduced recombinantly into a methylotrophic yeast as an additional copy despite the presence of a corresponding native nucleic acid sequence. For example, *P. pastoris* contains an endogenous nucleic acid encoding a Mxr1 transcriptional activator; an additional *P. pastoris* Mxr1 nucleic acid (e.g., introduced recombinantly into *P. pastoris*) or modifying the endogenous *P. pastoris* Mxr1 nucleic acid is considered exogenous.

Transcriptional activators, and nucleic acids encoding transcriptional activators (e.g., exogenous nucleic acids encoding transcriptional activators), are known in the art. For example, a transcriptional activator from *Pichia pastoris* is the Mxr1 sequence, but suitable transcriptional activators also can be found in *Hansenula polymorpha* (the Adr1 sequence; see, for example, GenBank Accession No. AEOI02000005, bases 858873 to 862352, for the nucleic acid sequence and GenBank Accession No. ESX01253 for the amino acid sequence) and *Candida boidinii* (the Trm1 sequence; see, for example, GenBank Accession No. AB365355 for the nucleic acid sequence and GenBank Accession No. BAF99700 for the amino acid sequence; the Trm2 sequence; see, for example, GenBank Accession No. AB548760 for the nucleic acid sequence and GenBank Accession No. BAJ07608 for the amino acid sequence). A representative *P. pastoris*

Mxr1 nucleic acid sequence can be found, for example, in GenBank Accession No. DQ395124, while a representative *P. pastoris* Mxr1 polypeptide sequence can be found, for example, in GenBank Accession No. ABD57365.

Transcriptional activators such as Mxr1 may be normally expressed at low levels. Therefore, it is desirable to place the exogenous nucleic acid (i.e., the transcriptional activator) under control of a promoter that is inducible. As used herein, "operably linked" means that a promoter or other expression element(s) are positioned relative to a nucleic acid coding sequence in such a way as to direct or regulate expression of the nucleic acid (e.g., in-frame).

There are a number of inducible promoters that can be used when genetically engineering methylotrophic yeast. For example, a methanol-inducible promoter, or a promoter element therefrom, can be used. Methanol inducible promoters are known in the art. For example, a commonly used methanol-inducible promoter from *P. pastoris* is the promoter, or a portion thereof, from the alcohol oxidase 1 (AOX1) gene, which is strongly transcribed in response to methanol. Other methanol-inducible promoters, or promoter elements therefrom, however, can be used, including, without limitation, the alcohol oxidase (AOD1) promoter from *Candida boidinii* (see, for example, GenBank Accession No. YSAAOD1A), the alcohol oxidase (MOX) promoter from *Hansenula polymorpha* (see, for example, GenBank Accession No. X02425), the MOD1 or MOD2 promoter from *Pichia methanolica* (see, for example, Raymond et al., 1998, Yeast, 14:11-23; and Nakagawa et al., 1999, Yeast, 15:1223-30), the DHAS promoter from *P. pastoris* (see, for example, GenBank Accession No. FJ752551) or a promoter element therefrom, the formaldehyde dehydrogenase (FLD1) promoter from *Pichia pastoris* (see, for example, GenBank Accession No. AF066054), or the PEX8 promoter from *P. pastoris* (see, for example, Kranthi et al., 2010, Yeast, 27:705-11). In some embodiments, the transcriptional activator is a Mit1 sequence from *Pichia pastoris* (see, for example, GenBank Accession No. CAY70887). All of these promoters are known to be induced by methanol.

A skilled artisan would understand that the recombinant nucleic acid molecule described herein can be stably integrated into the genome of the methylotrophic yeast cell, or can be extrachromosomally expressed from a replication-competent plasmid. Methods of achieving both are well known and routinely used in the art.

As demonstrated herein, the methanol-regulated transcriptional activators in *Pichia* can bind to the AOX1 promoter and act cooperatively with Mxr1 to activate transcription from the

AOX1 promoter. In some embodiments, two methanol-regulated transcriptional activators (e.g., Mxr1 and Mit1) can be operably linked to a methanol inducible promoter element.

A strain that includes a recombinant nucleic acid molecule as described herein can be used to regulate (e.g., overexpress) a second recombinant nucleic acid molecule in the methylotrophic yeast cell. A second recombinant nucleic acid molecule can include, for example, one or more heterologous nucleic acids encoding one or more polypeptides of interest. Similar to the exogenous nucleic acid encoding the transcriptional activator, a heterologous nucleic acid refers to any nucleic acid sequence that is not native to the genome or in the genome of an organism (e.g., a heterologous nucleic acid can be a nucleic acid from one microorganism (e.g., one genus or species of methylotrophic yeast) that is introduced into a different genus or species of methylotrophic yeast).

Simply by way of example, heterologous nucleic acids encoding the one or more polypeptides of interest can be the nucleic acids involved in the biosynthesis of a heme-co-factor. Exemplified herein are nucleic acids encoding the 8 different enzymes involved in heme biosynthesis as determined and annotated from the sequence of the *Pichia pastoris* genome. For example, heterologous nucleic acids encoding ALA synthase, ALA dehydratase, porphobilinogen deaminase, UPG III synthase, UPG III decarboxylase, CPG oxidase, PPG oxidase, and ferrochelatase can be expressed in the methylotrophic yeast strains described herein. For genetically engineering methylotrophic yeast to contain more than one heterologous nucleic acids (e.g., transgenes), a combination of methanol-inducible and constitutive promoters, or elements therefrom, can be combined to further increase the expression of such nucleic acids.

Previous studies in *Saccharomyces cerevisiae* identified ALA dehydratase and porphobilinogen deaminase as rate limiting enzymes in heme biosynthesis (see, for example, Hoffman et al., 2003, Biochem. Biophys. Res. Commun., 310(4):1247-53). However, heterologous expression of individual heme enzymes in *P. pastoris* from the glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter failed to overcome limitations associated with the expression of a recombinant protein containing a heme co-factor (see Krainer et al., 2015, Microb. Cell Fact., 13;14:4). As described herein, highly efficient expression of a recombinant heme containing protein in *P. pastoris* was achieved by co-expressing the entire heme biosynthetic pathway from methanol-inducible promoters, although it would be appreciated that one or more

of the genes involved in the heme biosynthetic pathway could be expressed from one or more constitutive promoters.

In addition to the enzymes involved in iron-co-factor biosynthesis, it would be understood that a nucleic acid encoding a member of the globin family of proteins (PF00042 in the Pfam database) including plant hemoglobins can be present. In the Examples herein, a nucleic acid encoding soybean leghemoglobin (LegH) is present. LegH is a protein that binds to the iron co-factor, heme, which results in a characteristic absorption at 415 nm and a distinct red color. The LegH protein (also known as LGB2) is naturally found in root nodules of soybean (see, for example, UniprotKB Accession No. P02236), and the nucleic acid sequence used herein was codon optimized for expression in *P. pastoris*. See, for example, WO 2014/110539 and WO 2014/110532.

Alternatively, a heterologous nucleic acid encoding a polypeptide of interest can be, for example and without limitation, a dehydrin, a phytase, a protease, a catalase, a lipase, a peroxidase, an amylase, a transglutaminase, an oxidoreductase, a transferase, a hydrolase, a lyase, an isomerase, a ligase, or an antibody against any such polypeptides. In other embodiments, a heterologous nucleic acid can encode one or more enzymes involved in the pathway for production of small molecules, such as ethanol, lactic acid, butanol, adipic acid, or succinic acid.

Similar to the exogenous nucleic acid encoding the transcriptional activator, the heterologous nucleic acid encoding a polypeptide of interest can be operably linked to an inducible promoter element (e.g., a methanol-inducible promoter element), or the heterologous nucleic acid encoding a polypeptide of interest can be operably linked to a constitutive promoter or constitutive promoter element. Inducible promoters and elements therefrom are discussed above. Constitutive promoters and constitutive promoter elements are known in the art. For example, a commonly used constitutive promoter from *P. pastoris* is the promoter, or a portion thereof, from the transcriptional elongation factor EF-1α gene (TEF1), which is strongly transcribed in a constitutive manner. Other constitutive promoters, or promoter elements therefrom, however, can be used, including, without limitation, the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter from *P. pastoris* (see, for example, GenBank Accession No. U62648.1), the promoter from the potential glycosyl phosphatidyl inositol (GPI)-anchored protein, GCW14p (PAS_chr1-4_0586), from *P. pastoris* (see, for example, GenBank Accession

No. XM_002490678), or the promoter from the 3-phosphoglycerate kinase gene (PGK1) from *P. pastoris* (see, for example, GenBank Accession No. AY288296).

Similar to the recombinant nucleic acid molecule described herein, the second recombinant nucleic acid molecule can be stably integrated into the genome of the methylotrophic yeast cell, or can be extrachromosomally expressed from a replication-competent plasmid.

It would be understood by the skilled artisan that a combination of inducible (e.g., methanol-inducible) and constitutive promoters (or promoter elements therefrom) can be combined to further increase the expression of any of the nucleic acids operably linked thereto.

It would be appreciated by a skilled artisan that a heterologous nucleic acid encoding a polypeptide of interest operably linked to a promoter element can be separate from the recombinant nucleic acid molecule described herein, or can be contiguous with the exogenous nucleic acid encoding a transcriptional activator operably linked to a promoter element contained within the recombinant nucleic acid molecule described herein. It also would be appreciated by a skilled artisan that, if the second nucleic acid molecule is contiguous with the recombinant nucleic acid molecule described herein, that a single promoter, or promoter element therefrom, can be used to drive transcription of both or all of the genes (e.g., the exogenous nucleic acid encoding the transcriptional activator as well as the one or more heterologous nucleic acids encoding the polypeptide(s) of interest).

Methods of introducing nucleic acids into methylotrophic yeast cells are known in the art, and include, without limitation, transduction, electroporation, biolistic particle delivery, and chemical transformation.

In addition, methods of culturing methylotrophic yeast cells are known in the art. See, for example, Pichia Protocols, *Methods In Molecular Biology*, 389, Cregg, Ed., 2007, 2nd Ed., Humana Press, Inc. Under some circumstances, it may be desirable to introduce or add methanol to the culture media, although, as demonstrated herein, methanol is not required to obtain efficient expression at high levels of one or more polypeptides of interest. Under some circumstances (e.g., when one or more nucleic acids encoding enzyme(s) involved in an iron-co-factor biosynthesis are expressed), it may be desirable to supplement the culture media with iron or a pharmaceutically or metabolically acceptable (or GRAS) salt thereof.

Pichia strains are able to grow on methanol as the sole carbon source. Methanol utilization is initiated by the conversion of methanol to formaldehyde by the action of alcohol

oxidase. The methylotrophic yeast, *Pichia pastoris*, contains two genes for alcohol oxidases, AOX1 and AOX2. Strains with reduced alcohol oxidase activity ("methanol utilization slow" or MutS strains) often produce more of a recombinant protein expressed from the AOX1 promoter than strains that do not have reduced alcohol oxidase activity. Strains mutated in both AOX genes and completely lacking alcohol oxidase activity cannot metabolize methanol, but can still be induced for expression from the AOX1 promoter by methanol. These strains retain the ability to use other carbon sources for growth, but still express heterologous proteins from the AOX1 promoter upon the addition of methanol. Because these strains do not metabolize methanol ("methanol utilization minus" or Mut- strains), much less methanol is required for induction of protein expression, and strains carrying these mutations avoid issues related to methanol feeding in large-scale fermentations. See, for example, Chiruvolu et al., 1997, Enzyme Microb. Technol., 21:277-83. It was determined herein that expression of LegH from the AOX1 promoter in Mut- strains greatly improved the LegH yield. Thus, a methylotrophic yeast having a mutation in both the AOX1 gene and the AOX2 gene can be used in the methods described herein.

The protein of interest, or a complex that includes one or more proteins of interest (e.g., heme-bound LegH, a dehydrin, a phytase, a protease a catalase, a lipase, a peroxidase, an amylase, a transglutaminase, an oxidoreductase, a transferase, a hydrolase, a lyase, an isomerase, a ligase, or an antibody) can be purified from the yeast cells. Methods of purifying polypeptides are known in the art. As used herein, a "purified" polypeptide is a polypeptide that has been separated or purified from cellular components that naturally accompany it. Typically, the polypeptide is considered "purified" when it is at least 70% (e.g., at least 75%, 80%, 85%, 90%, 95%, or 99%) by dry weight, free from the polypeptides and naturally occurring molecules with which it is naturally associated. Since a polypeptide that is chemically synthesized is, by nature, separated from the components that naturally accompany it, a synthetic polypeptide is "purified."

As used herein, nucleic acids can include DNA and RNA, and includes nucleic acids that contain one or more nucleotide analogs or backbone modifications. A nucleic acid can be single stranded or double stranded, which usually depends upon its intended use. Also provided are nucleic acids and polypeptides that differ from a given sequence. Nucleic acids and polypeptides can have at least 50% sequence identity (e.g., at least 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity) to a given nucleic acid or polypeptide sequence.

In calculating percent sequence identity, two sequences are aligned and the number of identical matches of nucleotides or amino acid residues between the two sequences is determined. The number of identical matches is divided by the length of the aligned region (i.e., the number of aligned nucleotides or amino acid residues) and multiplied by 100 to arrive at a percent sequence identity value. It will be appreciated that the length of the aligned region can be a portion of one or both sequences up to the full-length size of the shortest sequence. It also will be appreciated that a single sequence can align with more than one other sequence and hence, can have different percent sequence identity values over each aligned region.

The alignment of two or more sequences to determine percent sequence identity can be performed using the computer program ClustalW and default parameters, which allows alignments of nucleic acid or polypeptide sequences to be carried out across their entire length (global alignment). Chenna et al., 2003, Nucleic Acids Res., 31(13):3497-500. ClustalW calculates the best match between a query and one or more subject sequences, and aligns them so that identities, similarities and differences can be determined. Gaps of one or more residues can be inserted into a query sequence, a subject sequence, or both, to maximize sequence alignments. For fast pairwise alignment of nucleic acid sequences, the default parameters can be used (i.e., word size: 2; window size: 4; scoring method: percentage; number of top diagonals: 4; and gap penalty: 5); for an alignment of multiple nucleic acid sequences, the following parameters can be used: gap opening penalty: 10.0; gap extension penalty: 5.0; and weight transitions: yes. For fast pairwise alignment of polypeptide sequences, the following parameters can be used: word size: 1; window size: 5; scoring method: percentage; number of top diagonals: 5; and gap penalty: 3. For multiple alignment of polypeptide sequences, the following parameters can be used: weight matrix: blosum; gap opening penalty: 10.0; gap extension penalty: 0.05; hydrophilic gaps: on; hydrophilic residues: Gly, Pro, Ser, Asn, Asp, Gln, Glu, Arg, and Lys; and residue-specific gap penalties: on. ClustalW can be run, for example, at the Baylor College of Medicine Search Launcher website or at the European Bioinformatics Institute website on the World Wide Web.

Changes can be introduced into a nucleic acid molecule, thereby leading to changes in the amino acid sequence of the encoded polypeptide. For example, changes can be introduced into nucleic acid coding sequences using mutagenesis (e.g., site-directed mutagenesis, PCR-mediated mutagenesis) or by chemically synthesizing a nucleic acid molecule having such changes. Such nucleic acid changes can lead to conservative and/or non-conservative amino acid substitutions at

one or more amino acid residues. A "conservative amino acid substitution" is one in which one amino acid residue is replaced with a different amino acid residue having a similar side chain (see, for example, Dayhoff et al. (1978, in Atlas of Protein Sequence and Structure, 5(Suppl. 3):345-352), which provides frequency tables for amino acid substitutions), and a non-conservative substitution is one in which an amino acid residue is replaced with an amino acid residue that does not have a similar side chain. Nucleic acid and/or polypeptide sequences may be modified as described herein to improve one or more properties including, without limitation, increased expression (e.g., transcription and/or translation), tighter regulation, deregulation, loss of catabolite repression, modified specificity, secretion, thermostability, solvent stability, oxidative stability, protease resistance, catalytic activity, and/or color.

As used herein, an "isolated" nucleic acid molecule is a nucleic acid molecule that is free of sequences that naturally flank one or both ends of the nucleic acid in the genome of the organism from which the isolated nucleic acid molecule is derived (e.g., a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease digestion). Such an isolated nucleic acid molecule is generally introduced into a vector (e.g., a cloning vector, or an expression vector) for convenience of manipulation or to generate a fusion nucleic acid molecule, discussed in more detail below. In addition, an isolated nucleic acid molecule can include an engineered nucleic acid molecule such as a recombinant or a synthetic nucleic acid molecule.

Nucleic acids can be isolated using techniques routine in the art. For example, nucleic acids can be isolated using any method including, without limitation, recombinant nucleic acid technology, and/or the polymerase chain reaction (PCR). General PCR techniques are described, for example in PCR Primer: A Laboratory Manual, Dieffenbach & Dveksler, Eds., Cold Spring Harbor Laboratory Press, 1995. Recombinant nucleic acid techniques include, for example, restriction enzyme digestion and ligation, which can be used to isolate a nucleic acid. Isolated nucleic acids also can be chemically synthesized, either as a single nucleic acid molecule or as a series of oligonucleotides.

Polypeptides can be purified from natural sources (e.g., a biological sample) by known methods such as DEAE ion exchange, gel filtration, and hydroxyapatite chromatography. A polypeptide also can be purified, for example, by expressing a nucleic acid in an expression vector. In addition, a purified polypeptide can be obtained by chemical synthesis. The extent of

purity of a polypeptide can be measured using any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

A construct or vector containing a nucleic acid (e.g., a nucleic acid that encodes a polypeptide) also is provided. Constructs or vectors, including expression constructs or vectors, are commercially available or can be produced by recombinant DNA techniques routine in the art. A construct or vector containing a nucleic acid can have expression elements operably linked to such a nucleic acid, and further can include sequences such as those encoding a selectable marker (e.g., an antibiotic resistance gene). A construct or vector containing a nucleic acid can encode a chimeric or fusion polypeptide (i.e., a polypeptide operatively linked to a heterologous polypeptide, which can be at either the N-terminus or C-terminus of the polypeptide). Representative heterologous polypeptides are those that can be used in purification of the encoded polypeptide (e.g., 6xHis tag, glutathione S-transferase (GST))

Expression elements include nucleic acid sequences that direct and regulate expression of nucleic acid coding sequences. One example of an expression element is a promoter sequence. Expression elements also can include introns, enhancer sequences, response elements, or inducible elements that modulate expression of a nucleic acid. Expression elements can be of bacterial, yeast, insect, mammalian, or viral origin, and vectors can contain a combination of elements from different origins.

Vectors as described herein can be introduced into a host cell. As used herein, "host cell" refers to the particular cell into which the nucleic acid is introduced and also includes the progeny of such a cell that carry the vector. A host cell can be any prokaryotic or eukaryotic cell. For example, nucleic acids can be expressed in bacterial cells such as *E. coli*, or in insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art. Many methods for introducing nucleic acids into host cells, both *in vivo* and *in vitro*, are well known to those skilled in the art and include, without limitation, electroporation, calcium phosphate precipitation, polyethylene glycol (PEG) transformation, heat shock, lipofection, microinjection, and viral-mediated nucleic acid transfer.

Nucleic acids can be detected using any number of amplification techniques (see, e.g., PCR Primer: A Laboratory Manual, 1995, Dieffenbach & Dveksler, Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; and U.S. Patent Nos. 4,683,195; 4,683,202; 4,800,159; and 4,965,188) with an appropriate pair of oligonucleotides (e.g., primers). A number

of modifications to the original PCR have been developed and can be used to detect a nucleic acid.

Nucleic acids also can be detected using hybridization. Hybridization between nucleic acids is discussed in detail in Sambrook et al. (1989, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; Sections 7.37-7.57, 9.47-9.57, 11.7-11.8, and 11.45-11.57). Sambrook et al. discloses suitable Southern blot conditions for oligonucleotide probes less than about 100 nucleotides (Sections 11.45-11.46). The Tm between a sequence that is less than 100 nucleotides in length and a second sequence can be calculated using the formula provided in Section 11.46. Sambrook et al. additionally discloses Southern blot conditions for oligonucleotide probes greater than about 100 nucleotides (see Sections 9.47-9.54). The Tm between a sequence greater than 100 nucleotides in length and a second sequence can be calculated using the formula provided in Sections 9.50-9.51 of Sambrook et al.

The conditions under which membranes containing nucleic acids are prehybridized and hybridized, as well as the conditions under which membranes containing nucleic acids are washed to remove excess and non-specifically bound probe, can play a significant role in the stringency of the hybridization. Such hybridizations and washes can be performed, where appropriate, under moderate or high stringency conditions. For example, washing conditions can be made more stringent by decreasing the salt concentration in the wash solutions and/or by increasing the temperature at which the washes are performed. Simply by way of example, high stringency conditions typically include a wash of the membranes in 0.2X SSC at 65°C.

In addition, interpreting the amount of hybridization can be affected, for example, by the specific activity of the labeled oligonucleotide probe, by the number of probe-binding sites on the template nucleic acid to which the probe has hybridized, and by the amount of exposure of an autoradiograph or other detection medium. It will be readily appreciated by those of ordinary skill in the art that although any number of hybridization and washing conditions can be used to examine hybridization of a probe nucleic acid molecule to immobilized target nucleic acids, it is more important to examine hybridization of a probe to target nucleic acids under identical hybridization, washing, and exposure conditions. Preferably, the target nucleic acids are on the same membrane.

A nucleic acid molecule is deemed to hybridize to a nucleic acid but not to another nucleic acid if hybridization to a nucleic acid is at least 5-fold (e.g., at least 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 20-fold, 50-fold, or 100-fold) greater than hybridization to another nucleic acid. The amount of hybridization can be quantitated directly on a membrane or from an autoradiograph using, for example, a PhosphorImager or a Densitometer (Molecular Dynamics, Sunnyvale, CA).

Polypeptides can be detected using antibodies. Techniques for detecting polypeptides using antibodies include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. An antibody can be polyclonal or monoclonal. An antibody having specific binding affinity for a polypeptide can be generated using methods well known in the art. The antibody can be attached to a solid support such as a microtiter plate using methods known in the art. In the presence of a polypeptide, an antibody-polypeptide complex is formed.

Detection (e.g., of an amplification product, a hybridization complex, or a polypeptide) is usually accomplished using detectable labels. The term "label" is intended to encompass the use of direct labels as well as indirect labels. Detectable labels include enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials.

Methods are described herein that can be used to generate a strain that lacks sequences for selection (i.e., that lacks a selectable marker). These methods include using a circular plasmid DNA vector and a linear DNA sequence; the circular plasmid DNA vector contains a selection marker and an origin of DNA replication (also known as an autonomously replicating sequence (ARS)), and the linear DNA sequence contains sequences for integration into the *Pichia* genome by homologous recombination. The linear DNA molecule additionally can include nucleic acid sequences encoding one or more proteins of interest such as, without limitation, heme-bound LegH, a dehydrin, a phytase, a protease a catalase, a lipase, a peroxidase, an amylase, a transglutaminase, an oxidoreductase, a transferase, a hydrolase, a lyase, an isomerase, a ligase, one or more enzymes involved in the pathway for production of small molecules, such as ethanol, lactic acid, butanol, adipic acid or succinic acid, or an antibody against any such proteins.

Pichia cells can be transformed with both DNA molecules and the transformants selected by the presence of the selectable marker on the circular plasmid. Transformants then can be screened for integration of the linear DNA molecule into the genome using, for example, PCR.

Once transformants with the correct integration of the marker-free linear DNA molecule are identified, the cells can be grown in the absence of selection for the circular plasmid. Because the marker-bearing plasmid is not stably maintained in the absence of selection, the plasmid is lost, often very quickly, after selection is relaxed. The resulting strain carries the integrated linear DNA in the absence of heterologous sequences for selection. Therefore, this approach can be used to construct *Pichia* strains that lack a selectable marker (e.g., a heterologous selection marker) with little to no impact on recombinant protein yield.

In accordance with the present invention, there may be employed conventional molecular biology, microbiology, biochemical, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. The invention will be further described in the following examples, which do not limit the scope of the methods and compositions of matter described in the claims.

EXAMPLES

PART A. Materials and Methods

Example 1—Polymerase Chain Reaction

Genes of interest were amplified from genomic DNA or plasmid DNA templates using Phusion Hi-fidelity DNA polymerase (New England Biolabs). Briefly, 0.6 µM each of forward and reverse primers are incubated with 10-50 ng of template DNA and 400 µM of nucleotide mix in the presence of 1-2 U of Phusion DNA polymerase. The reaction conditions were as follows:

1 cycle	Initial Denaturation	98°C	1min
25 cycles	Denaturation	98°C	10 sec
	Annealing		20 sec
	Extension	72°C	30 sec per kb
1 cycle	Final Extension	72°C	5min
1 cycle	Hold	4°C	Forever

Example 2—Plasmid Construction by Ligation

50-100 ng of restriction enzyme digested plasmid and 3X molar excess of PCR amplified inserts were incubated in the presence of T4 DNA ligase (New England Biolabs). Ligation was

carried out at 16°C for greater than 2 hr. 2 µl of ligation reaction was transformed into DH10B electrocompetent *E. coli* cells

Example 3—Transformation into *E. coli* ElectroMax DH10B T1 Phage-Resistant Competent Cells

1.5-2 µl of ligation mixture was transformed into 20 µl of ElectroMax DH10B T1 Phage-Resistant Competent Cells (Invitrogen, Cat # 12033-015) by electroporation using MicroPulser (BioRad) set at 1.7 kV using a 1 mm gap cuvette (BioRad, Cat # 165-2089); after a pulse 1 ml SOC was added to cells and cells were incubated at 37°C for 1 h with shaking at 200 rpm. 10 µl of recovery mixture was plated on LB agar plates containing ampicillin at a concentration of 100 µg/ml. Plates were incubated overnight at 37°C.

Example 4—Linearization of Plasmid DNA for Transformation into *P. pastoris*

Plasmid DNA was digested with either PmeI restriction endonuclease (New England BioLabs, Cat # R0560L) in 1x CutSmart Buffer for 1-4 hours at 37°C or SfiI restriction endonuclease in 1x CutSmart Buffer for 1-4 hours at 50°C (New England BioLabs, Cat # R0123L). Linearized plasmid was gel purified from a 0.8% agarose gel using Zymoclean Gel DNA Recovery Kit (Zymo Research, Cat # D4002). DNA was eluted in 20 µl H_2O.

Example 5—Preparation of *P. pastoris* Transformation-Competent Cells

Selected strains of *P. pastoris* were grown to mid-exponential growth phase (~2 OD) in 25 ml YPD medium. Cells were collected by centrifugation at 930 x g for 15 minutes. The cell pellet was resuspended in 2 ml of a solution of 80% YPD and 200 mM HEPES, pH 6.8. 75 µl of 1 M DTT was added. The resuspended cell pellet was mixed at 100 rpm at 30°C for 25 minutes. A 40 ml volume of ice cold, sterile water was added to the suspension, and the cells were collected by centrifugation at 1125 x g for 15 minutes and placed on ice. The cell pellet was resuspended in 40 ml ice cold water and collected as before for two additional wash steps. The cell pellet was then resuspended in 20 ml of ice cold 1 M sorbitol and collected by centrifugation as before. The final cell pellet was suspended in 0.3 ml ice cold, sterile sorbitol, aliquoted and frozen at -80°C.

Example 6—Transformation into *P. pastoris*

30-100 ng of linearized plasmid DNA was transformed into 30 µl of electrocompetent *P. pastoris* cells using a 1 mm gap GenePulser cuvette (BioRad) with a GenePulser (BioRad) set at 1.15 kV. 1 ml of YPD/1M sorbitol was added and mixed at a 1:1 ratio to the cells. The cells were allowed to recover for 3 h at 30°C with shaking at 100 rpm. 100 µl of the recovery mixture was plated on a YPD plate containing the appropriate antibiotic, and the rest of the cells were plated on a YPD plate with the appropriate antibiotic. Plates were incubated at 30°C for 48 hours. Primary transformation plates were streaked onto YPD plates with appropriate antibiotic, and plates were incubated for 48 h at 30°C. Individual clones were patched onto YPD plates with antibiotics and the patches were used to do colony PCR or gDNA prep to confirm integration into the chromosome and to grow the strains in shake flasks for further analysis.

Example 7—Growing Cultures in Shake Flasks for Production of LegH

A strain from a fresh patch was inoculated into growth media BMGY (BMY supplemented with 0.75% glycerol) and grown overnight at 30°C with shaking at 200 rpm. The next day, expression of LegH was induced with methanol by diluting the ON culture with BMMY media (BMY+ 1% methanol) supplemented with 0.1 mM Ammonium Fe(III) citrate. The culture was grown to an OD600 of 0.5-0.7. Antifoam was added to a final concentration of 0.01%. The cultures were grown for 72 hours total; cultures were supplemented with methanol every 24 hours by adding 1/10 of shake flask volume of 10x BMMY media (BMY+10% methanol). Cells were harvested after 72 h of induced growth by centrifugation.

Example 8—Shake Flask Medium

BMY media was prepared by dissolving 10 g of yeast extract and 20 g soytone in 790 ml water. The mixture was sterilized by autoclaving, and cooled to room temperature. 100 ml 1 M Potassium Phosphate buffer (pH 6.0) and 100 ml 10X Yeast Nitrogen Base without amino acids (13.4 g of YNB powder per 100 mL; Sigma-Aldrich) was filter sterilized (0.2 µm pore size PES) and added to the media. No pH adjustment is required.

BMY Media Components

Component	Amount, per 1L
Yeast Extract	10 g
Soy peptone (BD)	20 g
Yeast Nitrogen Base without amino acids, 10X solution	100 mL (results in 13.4 g/L in BMY)
1M Potassium Phosphate Buffer, pH 6.0	100 mL

The following components were dissolved in water, and autoclaved to sterilize.

Low-Osmolarity Medium for Shake Flask

Component	Amount, g/L
Ammonium Sulfate	15.7
Potassium Phosphate Monobasic	9.4
Calcium Sulfate Dihydrate	0.43
Magnesium Sulfate Heptahydrate	11.7
Sodium Citrate Dihydrate	1.13

Example 9—Fermentation Medium and Feeds

The components indicated below were dissolved and the volume adjusted with water. The components were FCC food grade or equivalent. The medium was sterilized by autoclaving, by steaming in place, or with an equivalent.

Low-Osmolarity Medium with 95 g/L Glycerol for Fermentation

Component	Amount, g/L
Ammonium Sulfate	15.7
Potassium Phosphate Monobasic	9.4
Calcium Sulfate Dihydrate	0.43
Magnesium Sulfate Heptahydrate	11.7
Sodium Citrate Dihydrate	1.13
Glycerol, USP grade 99.7%	95

After sterilization, the medium was allowed to cool down to room temperature, and the following was added:

Component	Amount, mL/L
Trace Metals PTM1 Solution	2
Vitamin Solution	4
Sigma 204 antifoam or equivalent	1

Trace metals PTM1 solution is available as a powdered mix from Sunrise Science (Cat No. 4052-A-B-1L). Pouch A and pouch B were mixed in 950 mL water, and 5 mL sulfuric acid was added. Some precipitation is expected upon mixing; the mixture was filter sterilized (0.2 µm pore size PES) and stored at 4°C in the dark.

Vitamin solution recipe

Component	Amount, g/L
biotin	0.2
calcium pantothenate	1
folic acid	0.2
inositol	1
niacin	0.2
p-aminobenzoic acid	0.2
Pyridoxine hydrochloride	1
riboflavin	0.5
thiamine hydrochloride	1
B12	0.1

Alternatively, trace metals PTM1 can be made as follows:

Component	Amount
Cupric sulfate – 5H2O	6.0 g
Sodium iodide	0.08 g
Manganese sulfate – H2O	3.0 g
Sodium molybdate – 2H2O	0.2 g
Boric acid	0.02 g
Cobalt chloride	0.5 g
Zinc chloride	0.5 g
Ferrous sulfate – 7H2O	65.0 g
Biotin	0.2 g

Sulfuric acid	5.0 ml
Water	To a final volume of 1 L

The components are mixed together, filter sterilized and stored at room temperature. The glycerol feed mix was prepared by mixing 17.5 g of AmberFerm 4000 into 320 mL water and stirring to dissolve. The water-Amberferm mixture was added to 850 g of glycerol and mixed well by vigorous stirring. The feed mix was sterilized by autoclaving.

Glycerol feed solution

Component	Amount, g/L
USP grade glycerol	850
Water	320
Sensient AmberFerm4000 soy hydrolysate	17.5

The methanol feed was made using 99-100% methanol supplemented with 12 mL/L of PTM1 solution.

Example 10—Protocol for Lab-Scale High Oxygen Transfer Fermentation

Seed Shake Flask Protocol

In a aseptic biosafety hood, low-osmolarity medium and BMY were mixed in a 9:1 low-osmo:BMY ratio. Glycerol, at a concentration of 12.5 g/L, was added to the medium. USP food grade glycerol/glycerin (99.7% purity in a 50% v/v (63% w/w) glycerol/water solution) was used and autoclaved to sterilize. Sigma 204 or an equivalent antifoam was added to the medium at a concentration of 0.25 mL/L. Glycerol seed vials were retrieved, sprayed outside with 70% IPA or ethanol and thawed inside a biosafety hood at room temperature for about 5 min. Baffled shake flasks were inoculated with glycerol seed vials; 1 mL of inoculum vial were used for every 1 L of shake flask medium. Cultures were grown at 30ºC for 24 hours with shaking (200 RPM with a 1" throw). A ratio of between 1:10 and 1:5 of actual medium volume: nominal shake flask volume was used. 2.8 L nominal volume flask with 250 to 500 mL of medium were routinely use with success. The OD at 600 nm was measured after 24 hours of growth; if the OD was 15 or higher, the culture was used to inoculate a fermenter. If the OD was less than 15, the culture was grown

for 1-2 more hours before the OD was determined again. If an OD of 15 was not reached after 15 to 30 hours, the seed flask was considered to have failed.

Fermentation Protocol

The fermentation medium and feeds were prepared as described herein. The initial volume should be about 40% of the maximum fermenter volume, e.g., 4 L, if the maximum working volume of the fermenter is 10 L. This is because the process will approach the maximum working volume by the end of the fermentation. The fermenter is inoculated with shake flask seed at 10% inoculum-fermenter ratio, e.g. if 4 L of initial media are present in the fermenter, the fermenter is inoculated with about 0.4 L of shake flask seed. The total volume in the fermenter at this point is referred to T0 volume, e.g. 4.4 L in this representative example. Process controls include the following: 30ºC temperature; dissolved oxygen controlled by agitation-aeration cascade to maintain a 20% saturation set point; and pH controlled via addition of 28% NH_4OH, the set point will depend on the phase of the process.

Batch phase (from inoculation to depletion of glycerol, signaled by DO spike): Depending on the responsiveness of the PID control for dissolved oxygen, a strong DO spike or a fast drop in agitation-aeration rates or a combination of both may be observed when the cells deplete the glycerol present in the medium. Fed-batch phase is initiated when this occurs. The duration of the batch phase is approximately 20 hours, but up to 24 hours is considered acceptable. The pH set point is 5.0. The wet cell weight at the end of the batch phase will be approximately 220 g/L.

Fed-batch phase: glycerol feed is initiated to achieve 12-14 g/L/hr of neat glycerol based on T0 volume. The federate was maintained until approximately 350 g/L wet cell weight was reached, which should take about 7-10 hours. The pH set point is 5.0.

Transition phase: A sample is taken before beginning the transition phase. Methanol feed was initiated to achieve 1 g/L/hr of neat methanol, based on the T0 volume, until 1-2 g/L methanol concentration was reached in the fermentation broth. The methanol feed rate was adjusted during the remainder of fermentation so as to maintain a methanol concentration of 0.25-1 g/L in broth. Glycerol federate was reduced from 12-14 g/L/hr to 8-9 g/L/hr of neat glycerol, based on T0 volume, linearly over the course of 2 hours. Stepwise reduction in feed rate every 20 min would be acceptable as well. The pH set point was changed to 3.5, and the fermentation was allowed to naturally adjust to the new set point (i.e., with no addition of acid).

Production phase (from end of glycerol feed ramp-down to end of fermentation): The pH set point was 3.5. A methanol concentration of 0.25-1 g/L was maintained in the fermentation broth. The feed rate of the glycerol was maintained at 8-9 g/L/hr of neat glycerol, based on the T0 volume. Samples were taken approximately every 12 hours. Samples were spun at 4000 to 7000 RCF at 4ºC, and the supernatant decanted. The supernatant was saved in a separate tube. Pellets and 3 samples of 5 mL of supernatants at each time point were frozen at -80ºC. If a 15-20% DO during production is unable to be maintained, even at maximum aeration and agitation rates for the vessel, the glycerol feed rate can be lowered up to 5 g/L/hr of neat glycerol, based on the T0 volume. Fermentation ended 60 hours after inoculation. At 1000 L scale, the harvest process consisted of shutting down feeds and aeration, chilling the broth to 8ºC, and concentrating the paste using a sharples or disk stack centrifuge. Harvesting usually takes about 5-10 hours and does not incur a detectable loss of quality of the product. For lab scale, it is sufficient to collect, in addition to the 3x5 mL samples, an additional 50 mL sample at the end. Wet cell weight was >450 g/L, and spun pellets looked pink, as opposed to spun pellets from pre-induction samples, which looked more white. The color change of the broth from white to a more pronounced pink started following about 6-12 hours of induction.

PART B. Construction of Production Strains

Production Strain MXY0183

Example 11—Cloning Each Enzyme of Heme Biosynthesis Pathway into pGAN or pGAZ Integration Vector

pGAN (with the nat selection marker) and pGAZ (with the zeocin selection marker) were purchased from Biogrammatics, Inc (Carlsbad, CA). Each gene was placed under control of the AOX1 promoter, and the FDH terminator was placed immediately after the stop codon of each gene. The genes in the heme biosynthesis pathway were PCR amplified from wild type *P. pastoris* strain or subcloned from previous constructs.

The heme biosynthetic pathway, including the enzymes involved in the biosynthesis of heme, are shown in Figure 1. The intermediates produced during the biosynthesis of heme are shown in the boxes, and the enzyme catalyzing each step is shown to the right. The rate limiting enzymatic steps, as shown in *S. cerevisiae*, are shown with underlining.

ALA synthase, ALA dehydratase, UPGIII synthase, UPGIII decarboxylase, CPG oxidase and PPG oxidase genes were PCR amplified with primers containing sites for recognition by the restriction endonuclease, BsaI. Oligonucleotides were synthesized by ElimBiopharm.

Primer Designation	Gene	Sequence	SEQ ID NO:
MxO0187	ALAsynth_F	GAGGGTCTCGGATGGAGTTTGTCGCCCGTC	19
MxO0188	ALAsynth_R	GAGGGTCTCGATTACAATCTGACTCCTGATGAGG	20
MxO0189	ALAdehyd_F	GAGGGTCTCGGATGGTGCATAAGGCTGAATACTTG	21
MxO0190	ALAdehyd_R	GAGGGTCTCGATTATTCAGATAACCACTCCAGG	22
MxO0191	UroporSynth_F	GAGGGTCTCGGATGCCAAAAGCCATTCTTCTGAAG	23
MxO0192	UroporSynth_R	GAGGGTCTCGATTAGTGCACTTTTTGTATAGAC	24
MxO0193	UroporDecarb_F	GAGGGTCTCGGATGAGTAGATTTCCAGAACTGAAG	25
MxO0194	UroporDecarb_R	GAGGGTCTCGATTATTGAGATCCAATGCG	26
MxO0195	CoproOx_F	GAGGGTCTCGGATGGCCATCGACTCTGATATC	27
MxO0196	CoproOx_R	GAGGGTCTCGATTATACCCATTCAATAGGAT	28
MxO0197	ProtoporOx_F	GAGGGTCTCGGATGCTGAAAAGTCTTGCACCAAA	29
MxO0198	ProtoporOx_R	GAGGGTCTCGATTAAATGCCACTGAGGGTAGC	30

Phusion High-Fidelity DNA Polymerase (New England BioLabs, Cat # M0530L) was used to amplify genes from genomic DNA. PCR products were obtained and purified using DNA Clean&Concentrator-5 (Cat # D4004) and DNA was eluted in 25 µl of H_2O. Vector DNA, pGAZ and pGAN, and PCR products were digested with BsaI (New England BioLabs, Cat # R0535S) in 50 µl reaction volume at 37°C.

Linearized vectors and digested PCR products were purified from 0.8% agarose gel using Zymoclean Gel DNA Recovery Kit (Zymo Research Cat # D4002). DNA was eluted in 20 µl H_2O. Ligation reactions were set up in 10 µl at 16°C overnight using T4 DNA Ligase (New England BioLabs, Cat # M0202S).

PBD and ferrochelatase genes were subcloned from previously constructed plasmids: pJAZ_PBD was digested with BstBI(Bsp119I) (ThermoScientific, FD0124)and NotI (ThermoScientific, FD0596) in 1x Fast Digest buffer for 5 min at 37°C. pJAZ_Ferroch was digested with MfeI (MunI, ThermoScientific, FD0753) and NotI (ThermoScientific, FD0596) in 1x Fast Digest buffer for 5 min at 37°C.

Digested products were purified from 0.8% agarose gel using Zymoclean Gel DNA Recovery Kit (Zymo Research Cat # D4002). DNA was eluted in 20 µl H_2O.

Ligation reactions were set up in 10 µl reaction at 16°C overnight using T4 DNA Ligase (New England BioLabs, Cat # M0202S).

1.5 µl of ligation mixture was transformed into 20 µl of ElectroMax DH10B T1 Phage-Resistant Competent Cells (Invitrogen, Cat # 12033-015) by electroporation using MicroPulser (BioRad) set at 1.7 kV; cells were incubated at 37°C in 1 ml SOC for 1 h with shaking at 200 rpm. 10 µl of recovery mixture was plated on LB agar plates containing ampicillin at concentration 100 µg/ml. Plates were incubated overnight at 37°C. Colonies were screened by colony PCR for the presence of the insert. The sequences of the genes were confirmed.

Designation	Construct	Gene
pMx0308	pGAN-ALAsynth	ALA synthase
pMx0309	pGAN-ALAD	ALAD
pMx0310	pGAN-UPGIIIsyn	Uroporphyrinogene synthase
pMx0311	pGAN-UPGIIIdecarb	Uroporphyrinogene decarboxylase
pMx0312	pGAN-CPGoxi	CPG oxidase
pMx0313	pGAN-PPGoxi	Protoporphyrin oxidase
pMx0314	pGAZ-ALAsyn	ALA synthase
pMx0315	pGAZ-ALAD	ALAD
pMx0316	pGAZ-UPGIIIsyn	Uroporphyrinogene synthase
pMx0317	pGAZ-UPGIIIdecarboxilase	UPGIII decarboxilase
pMx0318	pGAZ-PPGoxidase	PPG oxidase
pMx0319	pGAZ-CPG oxidase	CPG oxidase
pMx0320	pGAN-PBD	PBD
pMx0321	pGAZ-PGC	PBD
pMx0322	pGAZ-Fc	Ferrochelatase
pMx0323	pGAN-Fc	Ferrochelatase

Example 12 —Assembling Heme Biosynthesis Genes on Plasmids for Integration into *P. pastoris* mutS Genome

The whole cassette "promoter-gene-terminator" was PCR amplified with primers containing sites for restriction endonucleases to assemble plasmids for integration into the *Pichia* genome.

Assembling Paox1_UPS_FDHterm- Paox1_UPD_FDHterm-Paox1_CPO_FDH term. on pGAN plasmid (pMx327)

pGAN-CPGoxidase (pMx312) was used as a vector to clone the UPS and UPD cassettes. UPG III synthase cassette was PCR amplified from pMx310 with primers to AOX1 promoter / FDH1 terminator containing NheI and SphI recognition sites for restriction endonucleases correspondingly:

Primer Designation	Recognition Sites	Sequence	SEQ ID NO:
MxO0399	NheI-pAOX1-F	CAA TCG CTA GCA TCC AAC ATC CAA AGA CGA AAG G	31
MxO0401	SphI-FDH1-R	GGA TAG CAT GCA CCT TAT CAA GAT AGC TAG AAA TAG AAA TGG	32

UPG III decarboxylase cassette was PCR amplified from pMx311 with primers to AOX1 promoter / FDH1 terminator containing SphI and AgeI recognition sites for restriction endonucleases correspondingly:

Primer Designation	Recognition Sites	Sequence	SEQ ID NO:
MxO0402	SphI-pAOX1I-F	CAA TAG CAT GCA ACA TCC AAA GAC GAA AGG TTG AAT G	33
MxO0404	AgeI-FDH1-R	CAT GGT ACC GGT ACC TTA TCA AGA TAG CTA GAA ATA GAA ATGG	34

Phusion High-Fidelity DNA Polymerase (New England BioLabs, Cat # M0530L) was used to amplify DNA from plasmids.

Obtained PCR products were purified using DNA Clean&Concentrator-5 (Zymo Research, Cat # D4004) and DNA was eluted in 25 µl of H_2O.

pGAN-CPGoxidase (pMx312) designated as a vector was digested in 1x CutSmart Buffer with NheI-HF (New England BioLabs, Cat # R3131S) and AgeI-HF (New England BioLabs, Cat # R3552S) over night at 37°C.

UPG III synthase cassette PCR product was digested in 1x CutSmart Buffer with NheI-HF (New England BioLabs, Cat # R3131S) and SphI-HF (New England BioLabs, Cat # R3182S) over night at 37°C.

UPG III decarboxylase cassette PCR product was digested in 1x CutSmart Buffer with SphI-HF (New England BioLabs, Cat # R3182S) and AgeI-HF (New England BioLabs, Cat # R3552S) over night at 37°C.

Digested vector and PCR products were gel purified from 0.8% agarose using Zymoclean Gel DNA Recovery Kit (Zymo Research, Cat # D4002). DNA was eluted in 20 µl H_2O.

Three-way ligation between UPG III synthase cassette digested with NheI-SphI, UPG III decarboxylase cassette digested with SphI-AgeI and a vector digested with NheI-AgeI was set up in 10 µl at 16°C overnight using T4 DNA Ligase (New England BioLabs, Cat # M0202S).

1.5 µl of ligation mixture was transformed into 20 µl of ElectroMax DH10B T1 Phage-Resistant Competent Cells (Invitrogen, Cat # 12033-015) by electroporation using MicroPulser (BioRad) set at 1.7 kV; cells were incubated at 37°C in 1 ml SOC for 1 h with shaking at 200 rpm. 10 µl of recovery mixture was plated on LB agar plates containing ampicillin at a concentration of 100 µg/ml. Plates were incubated overnight at 37°C. Colonies were screened by colony PCR for the presence of the insert. The sequences of the junctions between vector and inserts were confirmed.

Assembling Paox1_ALAsynthase_FDH1term.- Paox1_PPGoxidase_FDH1term- Paox1_Fc_ FDH1term.- Paox1_PBD_ FDH1term cassette (pMx330)

a. PCR amplification of gene-cassettes:

ALAsynthase cassette was PCR amplified from pMx310 with primers to AOX1 promoter / FDH1 terminator containing NheI and XhoI recognition sites for restriction endonucleases correspondingly:

Primer Designation	Recognition Sites	Sequence	SEQ ID NO:
MxO0399	NheI-pAOX1-F	CAA TCG CTA GCA TCC AAC ATC CAA AGA CGA AAG G	35
MxO0400	XhoI-FDH1-R	GAT ATT GCT CGA GAC CTT ATC AAG ATA GCT AGA AAT AGA AAT G	36

PPGoxidase cassette was PCR amplified from pMx313 with primers to AOX1 promoter / FDH1 terminator containing XhoI and AflII recognition sites for restriction endonucleases correspondingly:

Primer Designation	Recognition Sites	Sequence	SEQ ID NO:
MxO0403	XhoI-pAOX1-F	CAA TCT CGA GAA CAT CCA AAG ACG AAA GGT TG	37
MxO0437	AflII-FDH1-R	CAA CCA TTT CTA TTT CTA GCT ATC TTG ATA AGG TCT TAA GTC CA	38

Ferrochelatase cassette was PCR amplified from pMx323 with primers to AOX1 promoter / FDH1 terminator containing AflII and AgeI recognition sites for restriction endonucleases correspondingly:

Primer Designation	Recognition Sites	Sequence	SEQ ID NO:
MxO0404	AgeI-FDH1-R	CAT GGT ACC GGT ACC TTA TCA AGA TAG CTA GAA ATA GAA ATG G	39
MxO0436	AflII-pAOX1-F	TTA CTT AAG TCC AAC ATC CAA AGA CGA AAG GTT G	40

G418 marker was PCR amplified from pJAG plasmid purchased from Biogrammatics using the following primers:

Primer Designation	Recognition Sites	Sequence	SEQ ID NO:
MxO0438	Mlu-G418-F	TCA CAG ACG CGT TGA ATT GTC C	41
MxO0439	BbvCI-G418-R	TTG CTC CTC AGC TTA GAA GAA CTC GTC CAA CAT CAA GTG	42

Phusion High-Fidelity DNA Polymerase (New England BioLabs, Cat # M0530L) was used to amplify DNA from plasmids. The PCR products were obtained and purified using DNA Clean&Concentrator-5 (Zymo Research, Cat # D4004) and DNA was eluted in 25 µl of H_2O.

b. Preparation of vectors

pGAZ-PBD (pMx321) designated as a vector was digested in 1x CutSmart Buffer with NheI-HF (New England BioLabs, Cat # R3131S) and XhoI (New England BioLabs, Cat # R0146S) overnight at 37°C.

pGAZ-ALAsyn.-PBD (pMx328) was digested in 1x NEBuffer3.1 with MluI (New England BioLabs, Cat # R0198S) and BbvCI (New England BioLabs, Cat # R0601S) overnight at 37°C.

pGAG-ALAsyn-PBD (pMx332) was digested in 1x CutSmart Buffer with XhoI (New England BioLabs, Cat # R0146S) and AgeI-HF (New England BioLabs, Cat # R3552S) overnight at 37°C.

c. Making intermediate constructs and assembling a final cassette

Digested vector and PCR products were gel purified from 0.8% agarose using Zymoclean Gel DNA Recovery Kit (Zymo Research, Cat # D4002). DNA was eluted in 20 µl of H_2O.

pGAZ-PBD (pMx321) vector, digested with NheI-XhoI restriction endonucleases, was ligated with ALAsynthase cassette PCR product digested with the same enzymes in 10 µl reaction at 16°C overnight using T4 DNA Ligase (New England BioLabs, Cat # M0202S) to yield a pGAZ-ALAsyn.-PBD plasmid (pMx328).

pGAG-ALAsyn-PBD (pMx332) digested with XhoI and AgeI–HF restriction endonucleases was ligated with PPGoxidase cassette and Ferrochelatase cassette PCR products digested with XhoI, AflII and AflII, AgeI-HF restriction endonucleases correspondingly in a three-way ligation reaction using T4 DNA Ligase (New England BioLabs, Cat # M0202S) to yield a pGAG-ALAsynthase_PPGoxidase_Fc_PBD (pMx330).

1.5 µl of ligation mixture was transformed into 20 µl of ElectroMax DH10B T1 Phage-Resistant Competent Cells (Invitrogen, Cat # 12033-015) by electroporation using MicroPulser (BioRad) set at 1.7 kV; cells were incubated at 37°C in 1 ml SOC for 1 h with shaking at 200 rpm. 10 µl of recovery mixture was plated on LB agar plates containing ampicillin at a concentration of 100 µg/ml. Plates were incubated overnight at 37°C. Colonies were screened by colony PCR for the presence of the insert. The sequences of junctions between the vector and the inserts were confirmed.

Designation	Construct	Gene
pMx0327	pGAN-UPGsyn_UPGdecarb_CPGoxy	UPGsyn_UPGdecarb_CPGoxy
pMx0328	pGAZ-ALAsyn-PGC	ALAsyn-PBD
pMx0330	pGAG-ALAsyn_PBD_PPG_Fc	ALAsyn_PBD_PPG_Fc
pMx0332	pGAG-ALAsyn_PBD	G418 marker

Example 13—Integration of Linearized Plasmids with Gene Cassettes into *P. pastoris* Bg11 Genome

The plasmids that were used to generate the production strain, MXY0183, are shown in Figure 2. The steps taken to make the modifications that led to production strain, MXY0183, are depicted in Figure 3.

The first enzyme to be introduced into the *P. pastoris* Bg11 genome was ALAD. A plasmid containing pAOX1-driven ALAD (pMX229, Figure 2i and Figure 3) was linearized using PmeI restriction enzyme (New England BioLab). Linearized plasmid was purified from 0.8% agarose gel as described and transformed into *P. pastoris* using homologous recombination at the native AOX1 locus, generating strain MXY099 (Figure 3).

A plasmid containing two copies of the soybean LegH gene (sequence optimized for *P. pastoris*; SEQ ID NO:3) under the control of the pAOX1 promoter designated pMX282 (Figure 2ii and Figure 3) was linearized using SfiI restriction enzyme. Linearized plasmid was purified from a 0.8% agarose gel as described and transformed into the *P. pastoris* strain containing ALAD, generating the strain MXY0118 (Figure 3). qPCR was used and determined that strain MXY0118 contained several copies of the LegH gene, likely due to concatamerization of the plasmid, pMX282, at the time of recombination.

Plasmid pMX327 (Figure 2iii and Figure 3) containing genes encoding Uroporphyrinogen III synthase (UPS), Uroporphyrinogen III decarboxylase (UPD) and Coproporphyrinogen III oxidase (CPO) (the enzymes catalyzing steps 4, 5 and 6, respectively) under control of the AOX1 promoter was linearized with the SfiI restriction endonuclease and introduced into MXY0118, yielding strain MXY0170 (Figure 3).

Genes encoding ALA synthase (ALAS), Protoporphyrin III oxidase (PPO), Ferrochelatase (FC) and Porphobilinogen deaminase (PBD) (the enzymes catalyzing steps 1, 7, 8 and 3, respectively) from the *P. pastoris* genome were assembled on plasmid pMX330 (Figure 2iv and Figure 3). pMX330 was linearized with the SfiI restriction endonuclease and transformed into MXY0170, leading to the generation of strain MXY0183 (Figure 3). The genotype of MXY0183 was confirmed using PCR and qPCR.

Production Strain MXY0207

Example 14—Construction of pGAB Expression Vector

The pGAB expression vector (Figure 4A) was constructed by replacing the open reading frame of the Zeocin resistance gene in the pGAZ vector (BioGrammatics, Inc., Carlsbad, CA) with the open reading frame from the Blasticidin S deaminase (BSD) gene from *Aspergillus terreus*, which allows for selection of transformants carrying the plasmid with the antibiotic Blasticidin S.

The BSD open reading frame was amplified from a commercially synthesized DNA molecule using oligonucleotide primers MxO0476 and MxO0477 using a high fidelity polymerase chain reaction as described herein.

Primer Designation	Description	Sequence	SEQ ID NO:
MxO0477	BSD_Reverse	TTA GTC TTG CTC CTC AGC TTA GCC	43
MxO0476	BSD_Forward	TCA CAG ACG CGT TGA ATT GTC C	44

The BSD PCR product was purified by gel electrophoresis on a 1% agarose gel in 1x TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3) and visualized using SYBR Safe DNA gel stain (Life Technologies, Carlsbad, CA). The desired DNA fragment was excised from the agarose gel and the DNA was recovered using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA).

The purified BSD PCR product and pGAZ vector were digested with 10 units each of the MluI and BbvCI restriction endonucleases (New England Biolabs, Ipswich, MA) for 1 hour at 37°C in 1x NEBuffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA, pH 7.9 @ 25°C). Digested DNA products were recovered by gel electrophoresis as described above.

The purified, MluI and BbvCI digested BSD product and pGAZ vector were incubated with 400 units of T4 DNA ligase (New England Biolabs) in 1x T4 DNA ligase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5 @ 25°C) in a 20 µl reaction, at 16°C for 2 hours in a 20 µl reaction. Electrocompetent E. coli DH10B cells were transformed with 2 µl of the ligation reaction and antibiotic resistant transformants were selected on LSB agar plates supplemented with 100 µg/µl ampicillin.

Example 15—Construction of Mxr1 Expression Vector

The Mxr1 expression vector, pMx354, was constructed by introducing the Mxr1 open reading frame into the pGAB vector (Figure 4B). The Mxr1 open reading frame was inserted into pGAB with the translation start immediately downstream of the methanol-inducible alcohol oxidase 1 (AOX1) promoter from *Pichia pastoris* and the translation stop signal immediately followed by the transcription terminator sequence from the *P. pastoris* FDH1 gene.

The open reading frame encoding the Mxr1 protein was amplified from genomic DNA isolated from *Pichia pastoris* strain Bg11 MutS obtained from BioGrammatics, Inc. (Carlsbad, CA). The Mxr1 open reading frame was amplified from *P. pastoris* genomic DNA with primers MxO0495 (TTT TGC GGC CGC ATG AGC AAT CTA CCC CCA ACT TTT G (SEQ ID NO:45)) and MxO0496 (AAA AGC GGC CGC CTA GAC ACC ACC ATC TAG TCG GTT (SEQ ID NO:46)), which appended flanking NotI restriction endonuclease recognition sites. Amplification was accomplished using the polymerase chain reaction as described herein.

The amplified Mxr1 PCR product and the pGAB vector were digested with 10 units of NotI restriction endonuclease (New England Biolabs) for 1 hour at 37°C in 1x NEBuffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA, pH 7.9 @ 25°C). Following digestion, the NotI-digested pMx352 vector was treated with 5 units Antarctic phosphatase (New England Biolabs) for 15 minutes at 37°C in 1x Antarctic phosphatase buffer (50 mM Bis-Tris-Propane-HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, pH 6 @ 25°C).

The NotI-digested amplified Mxr1 fragment and pMx352 vector were separated by electrophoresis on a 1% agarose gel in 1x TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3) and visualized using SYBR Safe DNA gel stain (Life Technologies, Carlsbad, CA). The desired DNA fragments were excised from the agarose gel and the DNA was recovered using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA).

The NotI-digested fragment containing Mxr1 open reading frame was introduced into pGAB at a NotI site immediately downstream of the AOX1 promoter by ligation. A mixture containing 137 ng of NotI-digested DNA encoding the Mxr1 open reading frame and 60 ng of NotI-digested, phosphatase-treated pMx352 was incubate with 400 units of T4 DNA ligase (New England Biolabs) in 1x T4 DNA ligase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5 @ 25°C) in a 20 µl reaction, at 16°C, for 2 hours in a 20 µl reaction. Electrocompetent E. coli DH10B cells were transformed with 2 µl of the ligation reaction and

antibiotic resistant transformants were selected on LSB agar plates supplemented with 100 µg/µl ampicillin. Plates were incubated overnight at 37°C. Colonies were screened for the presence of the insert by PCR using primers MxO0495 and MxO0496. The sequence of the final vector was confirmed by DNA sequencing.

During cloning, 6 additional amino acids were introduced at the N-terminus of Mxr1. The Mxr1 open reading frame is shown under the section "Nucleic acid sequences", with residual amino acids from the cloning shown with underlining. *Pichia* production strains containing the Mxr1 sequence having the additional 6-amino acids at the N-terminus and *Pichia* strains containing the wild type Mxr1 (i.e., without the additional 6 amino acids at the N-terminus) were indistinguishable in fermentation tanks.

Example 16—Construction of Native Mxr1 Expression Vector

A plasmid containing the Mxr1 transcription regulator gene under the control of the pAOX1 promoter, designated pMX354, was used as a template for PCR amplification. The 3' end of the AOX1 promoter, the LegH open reading frame, and the AOX1 terminator were amplified from pMX354 using primers MxO0617 and MxO0647 shown below. The AOX1 terminator, linker and the 5' end of the AOX1 promoter were amplified from pMX382 using primers MxO0618 and MxO0646.

Primer Designation	Sequence	SEQ ID NO:
MxO0646	ACTAGATGGTGGTGTCTAGTCAAGAGGATGTCAGAATGCCATTTG	47
MxO0647	TCTGACATCCTCTTGACTAGACACCACCATCTAGTCGGTTTTCTAG	48

PCR products were obtained and purified using DNA Clean&Concentrator-5 and DNA was eluted in 12 µl of H_2O. The purified PCR products were then combined and used as a template for a subsequent round of PCR amplification using primers MxO0617 and MxO0618. The resulting PCR product was composed of the 3' end of the AOX1 promoter, followed by the Mxr1 open reading frame, the AOX1 terminator, a short linker sequence, and the 5' end of the AOX1 promoter. The PCR product was obtained and purified as described herein. The purified PCR product was cloned into the pCR™-Blunt II-TOPO® vector using the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen, Cat # K2800-20) to create the pMX402 vector.

Example 17—Construction of *P. pastoris* Strains MXY0206 and MXY0207

The pMx354 Mxr1 expression vector (Figure 4B) was introduced into the MXY0183 strain by DNA transformation (Figure 3).

The pMx354 vector (1.5 µg) was linearized at a unique PmeI site in the AOX1 promoter sequences by digestion with 20 units of the PmeI restriction endonuclease (New England Biolabs) for 1 hour at 37°C in 1x NEBuffer 4 (50mM Potassium Acetate, 20mM Tris-acetate, 10mM Magnesium Acetate, 1mM DTT, pH 7.9@25°C).

The PmeI-digested pMX354 vector was purified by gel electrophoresis recovered using the Zymoclean Gel DNA Recovery Kit as described above. The linearized pMX354 vector was introduced into strain MXY0183 by transformation and selection on blasticidin-containing medium. Two independent clones were obtained from the transformation, and they were designated MXY0206 and MXY0207. The presence of an additional copy of Mxr1 under the control of the AOX1 promoter in these strains was confirmed by PCR.

Production Strain MXY0291

Example 18—Construction of Strains MXY0213 and MXY0260

Figure 5 shows the steps taken to construct antibiotic marker free strain MXY0213 that contains 7 enzymes of the heme biosynthetic pathway. A linear piece of DNA containing variant Mxr1 (6 extra amino acids at N terminus) under pAOX1, with homology to the pAOX1 promoter on each end, was introduced using co-transformation (Figure 5 and Figure 6i). This linear Mxr1 expression cassette was simultaneously introduced into *Pichia* strain MXY213 with the pIL75 plasmid by transformation. The pIL75 vector carries a panARS autonomous replication sequence (Liachko & Dunham, 2014, FEMS Yeast Res., 14:364-7), which allows for maintenance of the plasmid vector without integration into the genome of the transformed cells, and a kanMX marker for selection of transformants with the antibiotic G418. Transformed cells were selected on media supplemented with G418 for the presence of kanMX marker on the pIL75 plasmid. *Pichia* transformants were screened by colony PCR for transformants that took up both the pIL75 plasmid and had correctly integrated the Mxr1 expression cassette.

Example 19—Co-Transformation to Introduce the LegH Expression Cassette into *Pichia*

A plasmid containing a different Pichia pastoris-codon optimized variant of soybean LegH gene (variant 3; SEQ ID NO:5) under the control of the pAOX1 promoter designated pMX399 was used as a source of template for PCR amplification of the gene. The backbone from TOPO cloning plasmid pMX401was PCR amplified. The insert and vector were assembled using Gibson assembly (NEB Gibson assembly kit) to generate plasmid pMX422. This plasmid was used as a template for a subsequent round of PCR amplification using primers MxO0617 and MxO0618 shown below.

Primer Designation	Sequence	SEQ ID NO:
MxO0617	AAACGCTGTCTTGGAACCTAATATGAC	49
MxO0618	AAACTGTCAGTTTTGGGCCATTTG	50

The resulting PCR product was composed of, in the 5' to 3' direction, the 3' end of the AOX1 promoter, followed by the LegH var 3 open reading frame, the AOX1 terminator, a short linker sequence, and the 5' end of the AOX1 promoter (Figure 6ii). The PCR product was obtained and purified by agarose gel electrophoresis as described herein.

Transformants with LegH expression cassette integrated into the genome were screened by PCR and characterized for LegH gene copy number using qPCR.

Example 20—Curing Transformants of Plasmid Vectors Bearing Selection Markers

In clones where the soybean LegH expression cassette was shown to be correctly integrated by colony PCR and in high copy number by qPCR, the pIL75 plasmid required for selection on G418 was eliminated by relaxing selection for the antibiotic. Transformants were streaked out for single colonies on media lacking G418 antibiotic. Because the panARS plasmid is not stably maintained in the absence of selection, the pIL75 was rapidly lost from the transformed cells under this condition. The resulting *Pichia* strain, MXY0291, contains sequences for LegH expression in copy number similar to MXY0207, but lacks heterologous sequences for selection.

Production Strains MXY0330, MXY0333, and MXY0338

Example 21—Construction of Strain MXY0306

Genotype PCR of strain MXY0291 revealed that a portion of the CPGoxidase coding sequence had been deleted during construction of this strain. The full-length CPGoxidase coding region was restored by replacement of the truncated copy. Briefly, a linear DNA fragment containing the pAOX1 promoter and full-length CPGoxidase coding region was generated by PCR amplification from plasmid pMX312 using primers MxO0866 and MxO0867 shown below.

Primer Designation	Sequence	SEQ ID NO:
MxO0866	ACGCTGTCTTGGAACCTAATATGAC	51
MxO0867	TACCCATTCAATAGGATTTTGTAGTACCTGC	52

The linear pAOX1-CPGoxidase DNA fragment was introduced into strain MXY0291 by co-transformation with the pIL75 plasmid. Transformants were selected on media containing G418 and then screened for the presence of the full-length CPGoxidase coding region by PCR. An isolate containing the full-length CPGoxidase was identified and subsequently cured of the plasmid vector required for selection on G418 as described above. This strain was designated MXY0306 (see Figure 5).

Example 22—Linear Constructs for Hybrid Promoter Strains

LegH variant 3 was expressed under the direction of each of the three native *Pichia pastoris* constitutive promoters indicated herein. The linear constructs are shown in Figure 7, and contained the 3' half of the promoter, followed by LegH var3, followed by the FDH1 transcription terminator. This was immediately followed by the antibiotic resistance cassette containing the pTEF promoter from *Ashbya gossypii*, the acetamidase gene (amdS) from *Aspergillus nidulans* and the TEF terminator from *Ashbya gossypii*. Finally, the construct contained the 5' half of the promoter. This linear cassette was amplified using the oligonucleotide primers listed in the Table below to generate constructs that contain several hundred base pairs on the 5' and 3' ends that are homologous to the respective promoter in the native *Pichia* genome.

Primers used to amplify the linear constructs

Primer designation	Sequence	SEQ ID NO:
MXO0718	GAGCTTCTTCTACGGCCCCC	53
MXO0723	TCCAGCAGAGTAAAATTTCCTAGGGAC	54
MXO0724	CTCTTTTAGGTTTTAAGTTGTGGGAACAGTAACA	55
MXO0729	GTGGGTGCTTCTTTGCGTGG	56
MXO0730	AGAATTGCCATCAAGAGACTCAGGACT	57
MXO0735	GATAGAGAGAAATCGCAAACTTTGAGAGGAAG	58

Competent MXY0306 cells were transformed with each of the linear cassettes and transformants containing the amdS selection cassette were selected based on their ability to grow on agar plates containing acetamide as the sole nitrogen source. These strains were purified, isolated and the presence of LegH under control of the constitutive promoter was verified by PCR (Figure 5).

Example 23—Nucleic Acid Sequences

Mxr1 nucleic acid sequence (the underlined nucleotides encode 6 amino acid at N-term introduced during cloning) (SEQ ID NO:1)

ATGCGAGACCGCGGCCGCATGAGCAATCTACCCCCAACTTTTGGTTCCACTAGACAATCTCCAGA
AGACCAATCACCTCCCGTGCCCAAGGAGCTGTCATTCAATGGGACCACACCCTCAGGAAAGCTAC
GCTTATTTGTCTGTCAGACATGTACTCGAGCATTTGCTCGTCAGGAACACTTGAAACGACACGAA
AGGTCTCACACCAAGGAGAAACCTTTCAGCTGCGGCATTTGTTCTCGTAAATTCAGCCGTCGAGA
TCTGTTATTGAGACATGCCCAAAAACTGCACAGCAACTGCTCTGATGCGGCCATAACAAGACTAA
GGCGCAAGGCAACTCGTCGGTCTTCTAATGCCGCGGGTTCCATATCTGGTTCTACTCCGGTGACA
ACGCCAAATACTATGGGTACGCCCGAAGATGGCGAGAAACGAAAAGTTCAGAAACTGGCCGGCCG
CCGGGACTCAAATGAACAGAAACTGCAACTGCAACAACAACATCTACAGCAACAACCACAGTTGC
AATACCAACAATCTCTTAAGCAGCATGAAAATCAAGTCCAGCAGCCTGATCAAGATCCATTGATA
TCCCCGAGAATGCAATTATTCAATGATTCCAACCATCACGTAAACAATTTGTTTGATCTTGGACT
AAGAAGAGCTTCCTTCTCCGCCGTTAGTGGAAATAATTATGCCCATTATGTGAATAATTTTCAAC
AAGATGCCTCTTCTACCAATCCAAATCAAGATTCAAATAATGCCGAATTTGAGAATATTGAATTT
TCTACCCCACAAATGATGCCCGTTGAAGATGCTGAAACTTGGATGAACAACATGGGTCCAATTCC
GAACTTCTCTCTCGATGTGAACAGGAACATTGGTGATAGCTTTACAGATATACAACACAAGAATT
CAGAGCCTATTATATCCGAACCGCCCAAGGACACCGCTCCAAACGACAAGAAGTTGAATGGCTAC
TCTTTTTACGAAGCCCCCATCAAGCCATTAGAATCCCTATTTTCTGTCAGGAATACAAAGAGAAA
CAAGTATAAAACAAATGACGACTCTCCAGACACCGTGGATAATAACTCCGCACCGGCTGCTAATA
CCATTCAAGAACTTGAGTCTTCTTTGAATGCATCCAAGAATTTTTGCTTGCCAACTGGTTATTCC
TTCTATGGTAATTTGGACCAACAGACTTTCTCTAACACGTTATCATGCACTTCTTCTAATGCCAC
AATTTCGCCCATTCTACTCGATAACTCCATTAATAATAACTCCACTAGTGACGTGAGACCAGAAT
TTAGAACACAAAGTGTCACCTCTGAAATGAGTCAAGCCCCTCCCCCTCCTCAAAAAAACAACTCG
AAATATTCCACCGAAGTTCTTTTTACCAGCAACATGCGGTCGTTTATTCACTACGCTCTTTCCAA
GTATCCTTTTATTGGTGTGCCCACTCCAACTCTTCCGGAGAACGAAAGACTAAATGAATATGCTG

ATTCATTCACCAACCGTTTCTTAAATCATTATCCTTTCATACATGTCACGATTCTCAAAGAATAC
TCCCTTTTCAAGGCAATTTTAGATGAGAATGAGTCGACTAAGAACTGGGAAAATAATCAGTTTTA
CTTAGAGAACCAACGAATATCAATTGTTTGTCTTCCTCTTTTGGTGGCTACGATAGGTGCAGTAC
TATCAAACAACAAAAAGGATGCTTCGAATTTATACGAAGCTTCAAGGCGTTGTATTCATGTTTAC
TTAGATTCCAGGAAAAAGATACCCACTTCCTTGTCCGCAAATAACAATGACTCTCCACTTTGGCT
AATTCAATCCCTGACGTTATCTGTTATGTATGGGTTATTTGCGGACAATGACATTAGTTTGAATG
TCGTGATCAGACAAGTTAACGCACTTAATTCTCTGGTCAAGACTTCGGGCCTGAATAGGACCTCA
ATTATAGATCTTTTCAACATCAACAAACCTTTGGATAATGAACTCTGGAATCAATTCGTGAAAAT
AGAGTCCACCGTAAGGACAATCCACACGATTTTTCAAATCAGTTCCAACTTAAGCGCCTTGTACA
ATATTATTCCATCGTTGAAAATTGATGACCTAATGATTACTCTACCAGTTCCCACAACACTTTGG
CAAGCTGATTCTTTTGTGAAATTCAAAAGTCTAAGTTACGGAAATCAGATCCCTTTTCAATATAC
AAGAGTACTACAGAATTTGATTGATTACAATCAGCCATTGAGCGATGGAAAATTTTTGTATGAAA
ACCATGTAAGTGAGTTTGGACTCATATGCCTACAGAATGGTCTACACCAATACAGCTATTTCCAA
AAATTGACTGCTGTCAATAACAGAGAAGATGCGCTATTCACAAAGGTTGTTAATTCACTTCACAG
TTGGGATAGGATGATTTCGAATTCTGATTTGTTTCCAAAGAAGATATATCAGCAGAGTTGCTTGA
TTTTGGACTCAAAGTTGCTTAATAATTTCCTGATTGTCAAGAGCTCATTGAAAGTTTCGACCGGA
GACGTTAGTTCTTTGAATAAGTTAAAAGAAAACGTGTGGCTTAAAAACTGGAATCAAGTGTGTGC
TATCTATTATAACAGCTTCATGAACATTCCTGCTCCCAGTATTCAAAAGAAGTACAATGACATAG
AGTTTGTGGATGACATGATTAATTTGAGTCTAATCATCATCAAGATTATGAAACTCATTTTCTAT
AACAATGTCAAAGACAATTATGAGGATGAAAATGACTTCAAATTGCAAGAGTTAAATTTAACATT
TGACAATTTTGATGAGAAAATATCCTTGAATTTGACAATATTATTCGATATATTTTTGATGATCT
ACAAGATAATTACCAATTACGAAAAGTTTATGAAGATCAAACACAAGTTTAATTACTACAATTCT
AATTCGAATATAAGCTTCTTGCATCATTTCGAACTCTCCTCGGTTATCAATAACACCCAAATGAA
CCAGAATGATTATATGAAAACAGATATTGATGAAAAGCTTGATCAGCTTTTCCACATCTATCAAA
CATTTTTCCGGCTGTATCTGGATTTAGAAAAGTTTATGAAGTTCAAATTCAACTATCATGACTTT
GAGACAGAGTTTTCAAGTCTCTCAATATCCAATATACTGAACACTCATGCTGCTTCTAACAATGA
CACAAATGCTGCTGATGCTATGAATGCCAAGGATGAAAAAATATCTCCCACAACTTTGAATAGCG
TATTACTTGCTGATGAAGGAAATGAAAATTCCGGTCGTAATAACGATTCAGACCGCCTGTTCATG
CTGAACGAGCTAATTAATTTTGAAGTAGGTTTGAAATTTCTCAAGATAGGTGAGTCATTTTTTGA
TTTCTTGTATGAGAATAACTACAAGTTCATCCACTTCAAAAACTTAAATGACGGAATGTTCCACA
TCAGGATATACCTAGAAAACCGACTAGATGGTGGTGTCTAG

Mxr1 protein sequence (the underlined 6 amino acid at N-term introduced during cloning) (SEQ ID NO:2)

MRDRGRMSNLPPTFGSTRQSPEDQSPPVPKELSFNGTTPSGKLRLFVCQTCTRAFARQEHLKRHE
RSHTKEKPFSCGICSRKFSRRDLLLRHAQKLHSNCSDAAITRLRRKATRRSSNAAGSISGSTPVT
TPNTMGTPEDGEKRKVQKLAGRRDSNEQKLQLQQQHLQQQPQLQYQQSLKQHENQVQQPDQDPLI
SPRMQLFNDSNHHVNNLFDLGLRRASFSAVSGNNYAHYVNNFQQDASSTNPNQDSNNAEFENIEF
STPQMMPVEDAETWMNNMGPIPNFSLDVNRNIGDSFTDIQHKNSEPIISEPPKDTAPNDKKLNGY
SFYEAPIKPLESLFSVRNTKRNKYKTNDDSPDTVDNNSAPAANTIQELESSLNASKNFCLPTGYS
FYGNLDQQTFSNTLSCTSSNATISPILLDNSINNNSTSDVRPEFRTQSVTSEMSQAPPPPQKNNS
KYSTEVLFTSNMRSFIHYALSKYPFIGVPTPTLPENERLNEYADSFTNRFLNHYPFIHVTILKEY
SLFKAILDENESTKNWENNQFYLENQRISIVCLPLLVATIGAVLSNNKKDASNLYEASRRCIHVY
LDSRKKIPTSLSANNNDSPLWLIQSLTLSVMYGLFADNDISLNVVIRQVNALNSLVKTSGLNRTS
IIDLFNINKPLDNELWNQFVKIESTVRTIHTIFQISSNLSALYNIIPSLKIDDLMITLPVPTTLW
QADSFVKFKSLSYGNQIPFQYTRVLQNLIDYNQPLSDGKFLYENHVSEFGLICLQNGLHQYSYFQ
KLTAVNNREDALFTKVVNSLHSWDRMISNSDLFPKKIYQQSCLILDSKLLNNFLIVKSSLKVSTG

DVSSLNKLKENVWLKNWNQVCAIYYNSFMNIPAPSIQKKYNDIEFVDDMINLSLIIIKIMKLIFY
NNVKDNYEDENDFKLQELNLTFDNFDEKISLNLTILFDIFLMIYKIITNYEKFMKIKHKFNYYNS
NSNISFLHHFELSSVINNTQMNQNDYMKTDIDEKLDQLFHIYQTFFRLYLDLEKFMKFKFNYHDF
ETEFSSLSISNILNTHAASNNDTNAADAMNAKDEKISPTTLNSVLLADEGNENSGRNNDSDRLFM
LNELINFEVGLKFLKIGESFFDFLYENNYKFIHFKNLNDGMFHIRIYLENRLDGGV*

Pichia pastoris-Codon-optimized LegH nucleic acid sequence (SEQ ID NO:3)
ATGGGTGCTTTCACCGAGAAGCAGGAAGCACTTGTTTCCTCTTCGTTCGAAGCTTTTAAGGCTAA
CATCCCTCAATACTCTGTTGTGTTTTACACGTCCATTCTAGAAAAAGCTCCTGCTGCCAAGGACC
TCTTCTCTTTTCTGTCCAACGGTGTAGATCCATCCAATCCCAAATTAACAGGTCACGCTGAGAAA
TTGTTCGGTTTAGTCAGAGATAGCGCTGGACAATTGAAAGCAAATGGTACTGTGGTTGCTGATGC
TGCCTTGGGCAGCATCCATGCACAGAAGGCAATTACAGACCCACAATTTGTTGTTGTGAAGGAAG
CTCTGCTTAAAACTATAAAGGAAGCCGTCGGAGACAAATGGAGTGACGAGTTGTCATCAGCTTGG
GAGGTAGCTTATGATGAGTTGGCCGCAGCAATCAAAAAGGCATTCTAA

Pichia pastoris-Codon-optimized LegH amino acid sequence (SEQ ID NO:4)
MGAFTEKQEALVSSSFEAFKANIPQYSVVFYTSILEKAPAAKDLFSFLSNGVDPSNPKLTGHAEK
LFGLVRDSAGQLKANGTVVADAALGSIHAQKAITDPQFVVVKEALLKTIKEAVGDKWSDELSSAW
EVAYDELAAA
IKKAF

Pichia pastoris-Codon-optimized LegH variant 3 nucleic acid sequence (SEQ ID NO:5)
ATGGGTGCATTTACAGAAAAACAAGAGGCTTTAGTATCCTCATCTTTTGAAGCTTTCAAAGCCAA
TATTCCTCAATACTCCGTTGTTTTCTATACGTCCATTTTGGAAAAGGCTCCAGCAGCTAAGGACC
TTTTCTCTTTCTTGTCGAACGGCGTGGATCCCTCAAATCCTAAGCTGACTGGTCACGCCGAGAAG
CTTTTTGGTTTGGTCAGAGACAGCGCCGGACAGCTGAAAGCTAACGGTACAGTTGTGGCAGATGC
TGCCTTGGGATCTATACATGCACAAAAGGCTATCACCGACCCACAGTTTGTGGTTGTAAAAGAGG
CTCTACTCAAAACTATCAAGGAAGCAGTTGGTGACAAATGGAGCGATGAATTGTCCAGTGCATGG
GAGGTCGCTTACGATGAGTTAGCTGCTGCAATCAAAAAGGCTTTCTAA

Pichia pastoris-Codon-optimized LegH variant 3 amino acid sequence (SEQ ID NO:6)
MGAFTEKQEALVSSSFEAFKANIPQYSVVFYTSILEKAPAAKDLFSFLSNGVDPSNPKLTGHAEK
LFGLVRDSAGQLKANGTVVADAALGSIHAQKAITDPQFVVVKEALLKTIKEAVGDKWSDELSSAW
EVAYDELAAAIKKAF

Pichia pastoris pAOX1 promoter (SEQ ID NO:7)
GATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCGACATCCACAGGTCCATT
TCTCACACATAAGTGCCAAACGCAACAGGAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAG
GACCTCCACTCCTCTTCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGC
TTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCATGACTTTATTAGCCT
GTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGTTTATTTCCGAATGCAACAAGCTCCGCATT
ACACCCGAACATCACTCCAGATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCA
AATGGCCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAGCGTGATCTCAT
CCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGCCAGTTGGTCAAAAAGAAACTTCCAA
AAGTCGGCATACCGTTTGTCTTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAAT
GCTTAGCGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAA
ACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGGGAA

TACTGCTGATAGCCTAACGTTCATGATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATA
TATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTACTTTC
ATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAA
GATCAAAAAACAACTAATTATTCGAAACG

Pichia pastoris pGAP promoter (SEQ ID NO:8)

CGACTATTATCGATCAATGAAATCCATCAAGATTGAAATCTTAAAATTGCCCCTTTCACTTGACA
GGATCCTTTTTTGTAGAAATGTCTTGGTGTCCTCGTCCAATCAGGTAGCCATCTCTGAAATATCT
GGCTCCGTTGCAACTCCGAACGACCTGCTGGCAACGTAAAATTCTCCGGGGTAAAACTTAAATGT
GGAGTAATGGAACCAGAAACGTCTCTTCCCTTCTCTCTCCTTCCACCGCCCGTTACCGTCCCTAG
GAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCCCTTGCAGCAATGCTCTTCCCAGCATTA
CGTTGCGGGTAAAACGGAGGTCGTGTACCCGACCTAGCAGCCCAGGGATGGAAAAGTCCCGGCCG
TCGCTGGCAATAATAGCGGGCGGACGCATGTCATGAGATTATTGGAAACCACCAGAATCGAATAT
AAAAGGCGAACACCTTTCCCAATTTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTATTT
GTCCCTATTTCAATCAATTGAACAACTATCAAAACACG

Pichia pastoris pGCW14 promoter (SEQ ID NO:9)

CAGGTGAACCCACCTAACTATTTTTAACTGGGATCCAGTGAGCTCGCTGGGTGAAAGCCAACCAT
CTTTTGTTTCGGGGAACCGTGCTCGCCCCGTAAAGTTAATTTTTTTTTCCCGCGCAGCTTTAATC
TTTCGGCAGAGAAGGCGTTTTCATCGTAGCGTGGGAACAGAATAATCAGTTCATGTGCTATACAG
GCACATGGCAGCAGTCACTATTTTGCTTTTTAACCTTAAAGTCGTTCATCAATCATTAACTGACC
AATCAGATTTTTTGCATTTGCCACTTATCTAAAAATACTTTTGTATCTCGCAGATACGTTCAGTG
GTTTCCAGGACAACACCCAAAAAAAGGTATCAATGCCACTAGGCAGTCGGTTTTATTTTTGGTCA
CCCACGCAAAGAAGCACCCACCTCTTTTAGGTTTTAAGTTGTGGGAACAGTAACACCGCCTAGAG
CTTCAGGAAAAACCAGTACCTGTGACCGCAATTCACCATGATGCAGAATGTTAATTTAAACGAGT
GCCAAATCAAGATTTCAACAGACAAATCAATCGATCCATAGTTACCCATTCCAGCCTTTTCGTCG
TCGAGCCTGCTTCATTCCTGCCTCAGGTGCATAACTTTGCATGAAAAGTCCAGATTAGGGCAGAT
TTTGAGTTTAAAATAGGAAATATAAACAAATATACCGCGAAAAAGGTTTGTTTATAGCTTTTCGC
CTGGTGCCGTACGGTATAAATACATACTCTCCTCCCCCCCCTGGTTCTCTTTTTCTTTTGTTACT
TACATTTTACCGTTCCGTCACTCGCTTCACTCAACAACAAAA

Pichia pastoris pTEF1 promoter (SEQ ID NO:10)

ATAACTGTCGCCTCTTTTATCTGCCGCACTGCATGAGGTGTCCCCTTAGTGGGAAAGAGTACTGA
GCCAACCCTGGAGGACAGCAAGGGAAAAATACCTACAACTTGCTTCATAATGGTCGTAAAAACAA
TCCTTGTCGGATATAAGTGTTGTAGACTGTCCCTTATCCTCTGCGATGTTCTTCCTCTCAAAGTT
TGCGATTTCTCTCTATCAGAATTGCCATCAAGAGACTCAGGACTAATTTCGCAGTCCCACACGCA
CTCGTACATGATTGGCTGAAATTTCCCTAAAGAATTTcTTTTTCACGAAAATTTTTTTTTTACAC
AAGATTTTCAGCAGATATAAAATGGAGAGCAGGACCTCCGCTGTGACTCTTCTTTTTTTTCTTTT
ATTCTCACTACATACATTTTAGTTATTCGCCAAC

Heme biosynthesis enzyme 1- ALA Synthase (SEQ ID NO:11)

ATGGAGTTTGTCGCCCGTCAGTCCATGAATGCCTGTCCCTTTGTCAGGTCAACTTCTACCCACCA
TTTGAAGAAGTTGGCAGCAAACAGTTCTCTAGCTGCTACTGCTAGTCATTGTCCCGTGGTTGGCC
CTGCTCTCCAACAGCAGAGATACTACTCTCAACCTTCCAAGCCAGCCCAAGCCCAAACCTCCGAC
ATTGCTACTGGGATCAAGAAGGATGTTTCTCCGATCCGTATGGACTCTAATGAAACCGCCTTTGA
TTACAATGGAATGTATGAGTCTGATCTTGCGAATAAACGTAAAGATAACTCGTATCGTTATTTCA

ATAACATCAACCGTCTAGCCAAGGAGTTTCCCAAGGCACATCGCCAGACCGAAGATGACAAGGTG
ACCGTCTGGTGCTCTAACGACTACTTAGGAATGGGTAGGCATCCTGAGATTATCAAAACCATGAA
GGCTACCATGGACAAGTACGGTTCCGGAGCAGGAGGAACTAGGAACATTGCAGGTCATAACCACG
CCGCTATCAATTTGGAAAGCGAGTTGGCTTGCTTGAACAAGAAGGAAGCGGCTCTGGTGTTTTCA
TCATGTTTCATAGCTAACGATGCAATCATCTCGTTGTTGGGACAAAAAATCAAAAATTTGGTCAT
TTTCTCTGACCAGTCGAATCATGCTTCCATGATATTGGGTGTGCGTAACTCCAAAGCGAAGAAGC
ACATCTTCAAGCACAACAATTTGAAGGATCTGGAGTCGCAGTTAGCTCAGTACCCCAAGTCGACT
CCTAAACTGATCGCCTTCGAGTCAGTTTACTCTATGTGTGGATCTGTGGCTCCCATTGAGAAGAT
TTGCGATTTGGCTAAAAGGTACGGTGCCCTCACCTTCTTGGATGAAGTTCATGCTGTTGGAATGT
ATGGTCCTCATGGACAGGGTGTAGCTGAGCATTTGGACTTTGATCTGCATTTACAGTCTGGAATC
GCCAGTCCTAGCGTGGTGGACAAACGCACCATATTGGATCGTGTCGACATGATTACTGGTACTTG
CGGAAAGTCATTTGGTACTGTTGGAGGTTACGTTGCTGGTAGTGCCAACCTAATTGATTGGTTAA
GATCCTATGCGCCAGGTTTCATTTTCACTACCACACTTCCTCCTGCTATCATGGCTGGTACAGCC
ACTTCTGTTCGTATTGTTAGGGCCGACATTGAGGCCCGTATCAAGCAACAGCTTAATACTCGCTA
CGTCAAAGACTCATTTGAAAACCTTGGTATTCCAGTCATTCCAAACCCAAGTCACATTGTTCCTG
TTCTAGTTGGAAATGCTGCAGATGCCAAGAAGGCATCCGATATGTTAATGAACAAACACCGTATT
TATGTTCAAGCTATTAACTACCCTACTGTGCCTGTCGGTGAAGAACGACTAAGGATTACTCCTAC
TCCAGGTCATGGAAAGGAGATTTGTGACCAGCTGATCAGCGCTGTCGACGATGTTTTTACTGAGC
TTAATTTACCAAGAATCAACAAATGGCAGTCCCAAGGTGGTCATTGCGGTGTTGGTGATGCTAAT
TACGTACCAGAACCCAATCTGTGGACTCAGGACCAGCTCAGCTTGACAAACCAAGACTTGCACTC
CAATGTGCACAACCCAGTGATTGAGCAGATCGAAACCTCATCAGGAGTCAGATTGTAG

Heme biosynthesis enzyme 2- ALA dehydratase (SEQ ID NO:12)

ATGGTGCATAAGGCTGAATACTTGGACGACCACCCAACTCAGATTTCCAGCATTCTTTCAGGAGG
TTACAACCACCCATTACTTCGTGAATGGCAACATGAACGTCAACTCAACAAAAACATGTTCATCT
TTCCCCTGTTTGTCACAGATCGACCAGACGAAGAAGAACTTATTCCTAGTCTACCTAATATCAAG
AGGTTTGGCGTTAACAAGTTGATTCCTTATGTAGGAGGTTTGGTTTCCAAAGGATTGAGGGCGGT
GATCCTATTTGGTGTTCCTCTGAAGCCCGGTGTGAAAGATGAAGAAGGAACGGCCGCTGATGATC
CAGAGGGACCTGTTATCCAAGCCATCAAACACTTGAGAAAGAACTTTCCTGACCTGTATATCATC
ACCGATGTCTGTCTATGTGAGTACACCAGCCATGGACATTGTGGAATACTATATGAGGATGGCAC
TATCAACAGAGAGCTCTCAGTCCGTCGTATTGCTGCTGTAGCTGTCAAATATGCTCAAGCTGGAG
CCAACTCTGTGGCTCCTTCTGATATGACTGACGGCAGAATAAGAGATATTAAAGAAGGCTTACTA
AGTGCAGGACTGGCACATAAAACGTTTGTTATGTCCTACGCTGCAAAATTCTCTGGTAATTTGTA
TGGCCCTTTCAGAGATGCTGCAGGTTCCTGTCCATCTCAAGGGGACAGAAAATGTTACCAGCTTC
CTTCTGGAGGAAAAGGGTTGGCCCATCGTGCTCTGATTCGTGATATGAATGAAGGCACTGATGGA
ATTATTGTCAAACCATCTACATTCTATTTGGACATTGTCGCTGATGCTTATCAGCTTTGTAAAGA
CTATCCTATCTGCTGTTACCAGGTTTCTGGAGAGTACGCCATGCTACATGCAGCGGCAGAGAAGA
ATATTGTTGATCTGAAATCAATCGCTTTTGAAGCTCATCAAGGATTCTTGCGGGCTGGAGCTCGT
TTAATCATTAGTTACTTTACCCCTGAATTCCTGGAGTGGTTATCTGAATGA

Heme biosynthesis enzyme 3- Porphobilinogen deaminase (SEQ ID NO:13)

ATGAACCAAATCGAACAGAGCGGACCCATTGATTGCAGTTCCTTGAAATTGGGGTCCCGAAAGTC
CGCTCTGGCTATAATCCAGGCAGAAATCGTCCGCCAATTGATATTGAAAGAATACCCTGAATTGG
AGACGAAGTTGGTCAGTGTGTCCACCCTGGGGGGACCAAGTCCAGAATAAAGCACTTTTCACGTTT
GGAGGAAAATCTTTGTGGACCAAAGAACTTGAGATGTTGTTGTTGGAGAGTGTGGGAGGATTTGA
CCAAATAGACATGATTGTACACTCGTTGAAAGACATGCCAACTCATTTACCAGACGAATTTGAGC
TGGGTTGCATTATTGAAAGAGAAGACCCTAGAGACGCTTTGGTCGTGCAAGATGGTTTATCTTAC

AAGTCATTGGCCGACCTTCCAGAGGGAGCTGTGGTCGGTACGTCTTCGGTTAGAAGATCGGCTCA
ACTACTGAAGAATTTCCCTCATCTGAAATTCAAATCTGTTAGAGGAAACCTTCAGACCAGACTAA
GAAAATTAGATGATCCAGATTCCGAGTACTGCTGTCTCCTCCTTGCAGCAGCCGGTTTAATCAGG
ACAGGCTTACAACACAGAATTTCAATGTATTTGAACGACGATGTGATGTACCACTCCGTCGGACA
AGGAGCATTAGGAGTAGAGATCAGAAAAGGTGACCAATTCATGAAAAATATCTGTGAAAAGATTG
GGCATAGAACCACCACCCTTCGTTGTCTTGCAGAGAGAGCACTGCTGAGATATCTAGAGGGAGGC
TGCTCGGTGCCAATTGGGGTCTCCACTATTTATAGCGAGGATACGAAGGAACTTACCATGAACTC
CCTAGTCGTCAGTTGTAACGGTCGTGACTCGGTAACAGAATCAATGACTGAAGTCGTGACTACTG
AAGAGCAAGCTGAAGATTTCGGTGAAAGGCTGGCCCAGAAGCTCATAGATCAAGGTGCGAAACGC
ATTCTTGACGAGATCAACTTCAACAAGATCAAAGAGATTAAGGAAGAGGGTTTACATTAA

Heme biosynthesis enzyme 4- Uroporphyrinogen III synthase (SEQ ID NO:14)

ATGCCAAAAGCCATTCTTCTGAAGAATAAAACTACACCGAAGGATCCTTATCTGGAGAACTTCGT
AAGTAGTGGCTACTCGACCGATTTCGTACCACTTTTAGATCATATTCACATGGAGAAATCTGAGA
TCATCGCATTTCTCAAGACTGACTACTTTTTGCATAAAACTTTGGCGTTTATTATTACGTCCCAA
AGAGCTGTAGAAATGCTGAATGAGTGTATGCAAATACTGAGACGTACTGATCCTGAAATTACACA
AATCATCTATAGTAAACCTGTCTATACAGTTGGCCCTGCCACCTACAGAATACTTGCGGATGCTG
GCTTCGTGGATCTACGAGGCGGAGATAAGGCAGGAAACGGATCCATTCTAGCCCAGATAATTTTG
AATGATGACATTTACACTGGAATTGAAGATTCTGACAAGCATATAACGTTTTTCACGGGAGAAAC
AAGGAGAGACATAATTCCCAAATGTTTACTCTCTAACAACTTTCAACTTTACGAAAAGATTGTCT
ACAAGACTCTTCCTAGGGATGATATCGTGACTAGATTCAAGTCTGCCGTTGACAGCATCGACCAA
TCGCAAAGAAGTTCCAGTTGGGTGGTCTTCTTTTCGCCTCAAGGAACAGAGGACATTGTAACGTA
TCTTCAACACACCAAAGACCAATTTAATATTGCATCTATCGGGCCAACCACAGAAAAATACCTTC
TAAGCAAAAACCTGAAACCAAAAGTTGTGGCACCTAAGCCAGAGCCTATCTCTTTACTATTGTCT
ATACAAAAAGTGCACTAA

Heme biosynthesis enzyme 5- Uroporphyrinogen III decarboxylase (SEQ ID NO:15)

ATGAGTAGATTTCCAGAACTGAAGAATGACCTTATTTTAAGGGCAGCTCGTGGTGAAAAAGTTGA
ACGTCCCCCAATATGGATTATGAGACAGGCCGGAAGATATCTTCCGGAGTACCATGAGGTCAAAG
GAGGTAGGGACTTCTTTGAAACTTGCAGGGATGCTGAGATTGCTTCTGAAATTACTATCCAGCCG
ATTACGCATTTTGACGGTCTGATCGATGCAGCTATTATCTTCAGTGATATCTTGGTGATTCCTCA
AGCTATGGGCATGGAAGTTAAGATGGTGGACAAAGTTGGCCCACAGTTCCCCAATCCGCTAAGAA
AACCGTCTGACTTGGATCATTTGAAAAAAGACGTTGACGTTTTGAAGGAACTCGATTGGGCCTTC
AAAGCTATCTCATTGACCAGAAAAAAACTCAATGGACGAGTGCCTTTGCTTGGATTTTGTGGTGC
TCCTTGGACTCTACTGGTTTATATGACTGAAGGAGGCGGTACCAAGATGTTTCGATTTGCAAAAG
AGTGGATCTACAAGTTTACCAAGGAATCTCATCAATTACTCCAACAGATCACTGACGTTGCAGTT
GAATTCTTAGCTCAGCAAGTTGTTGCAGGTGCCCAAATGTTACAAGTTTTTGAATCTTGGGGCGG
TGAATTGGGGCCTGATGAATTCGATGAGTTTTCTTTGCCTTATTTGAGACAGATTTCCTCTAAAC
TTCCCCTGAGGTTGAAGGAACTTGGAATCACAGAGAATGTTCCCATAACTGTCTTTGCTAAAGGC
TCTTGGTACGCCTTGGAGCAATTGTGCGACAGTGGTTATGATGTTGTCTCGTTGGATTGGTTATT
CCGTCCAAGTGATGCTGTCCAGATTGCTAACGGAAGAATCGCATTGCAAGGTAATCTTGACCCTG
GAACCATGTACGGCTCCAAAGAAACCATTTCCAAGAAAGTGGACAAAATGATCAAGGGTTTTGGT
GGAGGAAAGCAAAACTACATAATTAATTTTGGACACGGCACTCATCCATTCATGGATCCAGAACA
GATCAGATGGTTCTTACAAGAATGTCATCGCATTGGATCTCAATAG

Heme biosynthesis enzyme 6- Coproporphyrinogen III oxidase (SEQ ID NO:16)

ATGGCCATCGACTCTGATATCAATCTAAGCTCTCCCAATGATTCCATCCGTCAAAGGATGTTCGA
GCTTATCCAGCGGAAGCAACTCGAAATTGTCGCTGCATTGGAGGCAATTGAAGGAAACGATACCA
AATTTCGTTCTGATTCTTGGGAAAGAGGAGCCGAAGGTGGAGGAGGAAGATCTATGCTTATTCAA
GATGGAAGAGTGTTTGAAAAGGCTGGTGTAAATATTTCCAAGGTTCATGGCGTATTGCCTCCTCA
AGCTGTGAGCCAGATGAGAAATGACCACTCCAAGCTAGATCTGCCTGCGGGAACCTCTCTGAAGT
TCTTTGCCTGTGGGCTTTCGTTGGTCATTCATCCCCATAATCCCCATGCTCCAACTACCCATCTG
AATTATCGCTACTTCGAAACTTGGGATGAAACTGGAAAGCCTCACACCTGGTGGTTTGGGGGCGG
TGCTGATTTAACGCCTTCGTACCTGTATCCCGAGGATGCCAAGCAATTCCATCAAGCCCATAAGG
ATGCCCTGGACAAACACGATGTTAGCTTGTACCCGAGATTCAAAAAGTGGTGTGATGAATACTTT
CTGATCAAACATCGAAATGAAACTAGAGGTATTGGGGGTATTTTCTTTGATGATTTTGACGAGTT
TGATGCTGAGAGGTCCCTGAAGTTGGTTGAAGATTGTTTCAATGCTTTCTTGGAATCTTATCCCG
CTATCACTCGAAAAAGGATGGACACCCCTTCAACTGATGCTGAGAAGAACTGGCAACAAATTAGA
AGAGGAAGATATGTCGAATTCAACTTAGTATTGGATAGAGGTACTCAATTTGGTTTGAGAACGCC
TGGATCTCGTGTTGAAAGTATTTTGATGTCGTTGCCAAGAACAGCTGGTTGGGTCTATGATCATC
ATCCAGAGCCTGGCTCCAGAGAAGAGGAGTTATTGCAGGTACTACAAAATCCTATTGAATGGGTA
TGA

Heme biosynthesis enzyme 7- Protoporphyrinogen oxidase (SEQ ID NO:17)

ATGCTGAAAAGTCTTGCACCAAATTCCTCAATTGCCGTTTTAGGTTCAGGGATATCTGGATTGAC
TTTCAGCTTTTTTTTGAATCGGTTGCGTCCCGATGTTAAGATCCATATCTTTGAAAAATCCAAGC
AGGTTGGAGGATGGATCAGATCAGAAGAGCATGAAACTTTTCATTTTGAAAAGGGACCCAGAACT
TTGAGAGGCACAAATACGGGTACCTTGATGTTGTTGGATCTTCTTACCAAGATAGGAGCAAATGA
CAAGGTCCTGGGACTGCACAAAGATTCTCTTGCTAATAAAAAGTATCTGTTGTCCCCGTTCTCAG
ATGTTCACGGAAACAACGCAAAGCTTCTTCAAGTGCCACAGGATTTCAGCTCTTTTGTAAAGTTC
ATGTTTGACCCGTTGTCTAAGGATCTCATTCTCGGTCTTTTGAAAGAACCATGGCAACCAAAATT
AAAGTATTCAGATGAGTCGGTTGACCATTTTTTCAACAGAAGATTTGCTACCAAACTATCAGAGA
ATATCGTCAGCGCAATTGTGCATGGAATCTATGCGGGCGACGTGAAGAAGTTAAGTGTGAAAGCC
ATCTTCCCTAGGCTCCCTGAGATGGAACAGGAAAGTGGCTCTATTATAAGGTATATGATCGCCCA
ATACAGGACAAAAAAGAACGTCAAACAAAAAGTTGACCCTTTTTTGGCAGATTATGAAAAATTGA
TCGGTACATCTTTGAGTTTCAAAAATATTTCTTTGTTTCTGAAAAACTTTCCCATGCTGAGTTTT
CAGGGTGGACTACAGAAACTTCCCATCTCATTGAAGAACCATTTATCACAGATTGAAAACATCAA
GTTTCATTTTGACAGCAAAATCAAAAACATTGCTTTGGAGAGCGGTAAGGTGGCATTGACTGACC
ATGATCAGGTTTATCTTGTTGACCATGTGAGATCTACCATTAATACCAACGAATTGGCCAAAATC
ATTTCACCCGTTGTTCCAAGTTCTACTAAGAAAAAATCCGTTTTCAAATCCAAAGCGAATGGCCC
AGGGCTGGTCAAATGTTTGAGCTGGCTACACTATACAAATATACTAATGTGCAACATTTATATAC
CTAAGCACGTCTCAAAATCTATCACCGGATTTGGATACTTGGTTCCTCGATCAATGTCTTCTCAG
GCATCCAAACTTCTCGGTGTCATATTTGACTCAGACATCGAGACTGCAATGACTCCTAATTTTAC
AGAGGCCAACATTACGGCGATAAACAGTAACTCTGCATCTCCCAAGCAACTCCAAAAGTTTTCTG
ACCAATTCGTCAATAATGATCTCCCTAAATACACCAAGTTGACGCTAATGCTTGGAGGTCATTAT
CTCAAGTCGGAGGCAGACATGCCCGGTTCCGCAGAGAGTAAACATGCTGTCAAGGCGATTCTGTC
AAATCACCTGAATATTGATCTAGATGAGTTTGCATCTTTGCCAGACTTCAAGATGGAAATCACCA
AGATCCCCAACTGCATTCCCCAATATGAAGTTGGGTATCTTGATCTCAAGAGAAAGGTTCAGAAT
GCAGCCTCCAAAGAGTTCAACGACCAAATAAGTTTTGGAGGCATGGCATTTGGTGATGGTGTGGG
GATCCCTGACTGTGTCCAGAATGCATTCAAAGATTCGGCTACCCTCAGTGGCATTTAA

Heme biosynthesis enzyme 8- Ferrochelatase (SEQ ID NO:18)

ATGCTTAACCGTCGTTTCCAATCTACCGTGTCCTCGAGTCTGAACAAGGGCACTGGAATAGTGTT
CATGAATATGGGTGGTCCCTCCACTGTCAAGGAAACCTATGACTTTTTATTTCGTCTTTTCTCGG
ACGGAGATTTAATCCCGTTTGGCAGATTTCAGAACATCCTGGCCCGCTTCATTGCAAGTAGAAGA
ACACCCAAAATTGAATCCTACTACAAAGCTATCGGAGGTGGGTCTCCTATCCGAAAGTGGTCTGA
ATACCAGAGTTCTAAACTATGTGAAAAATTAGACATTATCAGTCCACAATCGGCTCCTCATAAGC
CTTATGTTGCCTTCAGATACGCTAATCCTCTCACTGAAGATACTTTACAAAAGATGAAAAATGAT
GGAATTACTAAGGCCATTGCCTTTTCTCAATATCCGCAATTTAGTTATTCAACCACCGGATCATC
GATTAACGAACTTTACAGGCAATCGAAAATTTTGGACCCTGATCAATCTATTAAATGGACAGTTA
TAGATCGCTGGCCTGACCACCCAGCCTTAGTTAAAACTTTCGCAGCTCATATCAAAGATACTCTA
AACAGATTCAAAACTGAAAATGGACTGACTGACACAAAAGACGTCGTCCTCCAATTCAGTGCTCA
TTCTTTACCAATGGATATTGTCAATAAAGGAGATTCGTATCCTGCAGAAGTCGCAGCGAGTGTCT
TTGCCATTATGAAAGAACTTAACTTCTCAAATCCTTATAAATTAACCTGGCAATCACAGGTTGGC
CCAAAGCCTTGGCTGGGTGCTCAAACTGAAAAAATTACCAAGCAGCTAGCATCCAGTGATGTTCC
TGGAGTCGTTTTGGTTCCTATTGCCTTTACCTCTGATCATATTGAAACTCTCCATGAACTGGATA
TTGAACTGATTCAAGAACTACCTAATCCTTCAAAAGTAAAGCGAGTTGAATCGTTGAACGGAGAC
CAAACTTTCATTGACTCCTTGGCAGAACTAGTGAAGAGTCACATTGATTCGAAGGTTGTATTTTC
CAACCAGTTGCCATTGGATTCCATGCTGGGAGTAGTGTCAGATAATTCCCTCACAGATCCAAAAG
AGTTTTTCAGAGCCCATTGA

Part C. Results and Discussion

Example 24— Characterization of Strain MXY0183

Optimum growth conditions for Strain MXY0183 include a target pH of 3.0 to 6.0 and temperatures of 28-35°C. In order to produce the LegH protein, strain MXY0183 must be alive and growing aerobically for a period of 6 days.

Expression of the genes associated with strain MXY0183 resulted in phenotypic changes to the strain. Figure 8 shows photographs of shake flasks at the start of induction (0 hr) and 72 hr post-induction. The flasks designated #1 contains the host strain, MXY0051. The flasks designated #2 and #3 contain one of the intermediate strains (i.e., MXY0118, containing >10 copies of the LegH gene and the ALA dehydratase from the heme biosynthetic pathway) and the production strain (i.e., MXY0183, containing > 10 copies of the LegH gene and the 8 enzymes from the heme biosynthetic pathway), respectively. The characteristic red color in flask #3 after 72 hours demonstrates the production of heme-bound LegH.

After growing in shake flasks, the *P. pastoris* strains indicated above, MXY0051, MXY0118, and MXY0183, were lysed and the proteins run on a SDS gel (Figure 9A). The arrow shows the position of the LegH protein. A comparison of LegH production in strain MXY0183

and in strain MXY0118 is shown in Figure 9B, which demonstrates the efficiency of heme loading of the LegH protein by the MXY0183 strain.

Example 25— Characterization of Strain MXY0207

Experiments were then performed to determine the benefits of overexpressing the transcriptional activator, Mxr1, in the presence of the genes encoding the 8 enzymes involved in heme biosynthesis. Strain MXY0183, which contains >10 copies of the LegH sequence and the genes encoding the 8 enzymes involved in heme biosynthesis, and sister strains MXY0206 and MXY0207, which contain >10 copies of the LegH sequence, the genes encoding the 8 enzymes involved in heme biosynthesis, and the Mxr1 transcriptional activator, were grown in shake flask cultures in the presence of glycerol, which is a repressing carbon source for these strains. Photographs of the shake flask cultures after 48 hr are shown in Figure 10A, and photographs of the pellets from cells grown on BMY media for 48 hours with no additional source of carbon are shown in Figure 10B; these experiments demonstrated that significant expression of transgenes (e.g. heme enzymes) under the control of the AOX1 promoter occurs in the absence of an inducing carbon source when a repressing carbon source is consumed in the growth medium of a strain in which Mxr1 is also expressed from the AOX1 promoter. The relative yield of heme-loaded LegH, when shake flask cultures were grown in the absence of induction agent, is shown in Figure 10C. These experiments demonstrated that significant production of a recombinant, heme-loaded protein is accomplished from AOX1 promoter-driven transgenes in the absence of methanol induction in Pichia strains in which Mxr1 expression is also driven by the AOX1 promoter.

Select strains were grown in 2 L fermenter tanks, and the relative yield of LegH and heme-loaded LegH was determined (Figure 11). Compared to strain MXY0183, the MXY0207 strain produced even more LegH and was able to produce enough heme to heme-load the LegH protein very effectively.

Example 26—Characterization of Strain MXY0291

As described above in Examples 18-20, strain MXY0291 was constructed to recapitulate the LegH production ability of MXY0207, while being free of antibiotic resistance genes. It was determined that strain MXY0291 contained ~ 16 copies of LegH var3, Mxr1 and 7 of the 8 heme

biosynthetic enzymes. When grown in 2 L fermenter tanks, this strain showed improved LegH yield compared to MXY0207. This improvement was seen both in induction media containing methanol / glycerol and methanol / dextrose (D-glucose) (Figure 11).

Example 27—Characterization of Hybrid Promoter Strains

Additional copies of soybean leghemoglobin (LegH) were expressed under three different constitutive promoters, pGAP, pGCW14 and pTEF1, in a strain that already contains several copies of LegH, all heme biosynthetic enzymes, and the transcriptional factor Mxr1 under control of the promoter, pAOX1 (referred to above as MXY0291). When induced by methanol in the presence of dextrose (i.e., D-glucose), the constitutive promoters and pAOX1 drive expression of LegH while only the pAOX1 promoter drives expression of the heme enzymes. This leads to further improvement in LegH yield compared to previous strain MXY0291 (Figure 11).

It is to be understood that, while the methods and compositions of matter have been described herein in conjunction with a number of different aspects, the foregoing description of the various aspects is intended to illustrate and not limit the scope of the methods and compositions of matter. Other aspects, advantages, and modifications are within the scope of the following claims.

Disclosed are methods and compositions that can be used for, can be used in conjunction with, can be used in preparation for, or are products of the disclosed methods and compositions. These and other materials are disclosed herein, and it is understood that combinations, subsets, interactions, groups, etc. of these methods and compositions are disclosed. That is, while specific reference to each various individual and collective combinations and permutations of these compositions and methods may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a particular composition of matter or a particular method is disclosed and discussed and a number of compositions or methods are discussed, each and every combination and permutation of the compositions and the methods are specifically contemplated unless specifically indicated to the contrary. Likewise, any subset or combination of these is also specifically contemplated and disclosed.

WHAT IS CLAIMED IS:

1. A methylotrophic Pichia yeast cell comprising:

a nucleic acid molecule encoding a heme-containing protein operably linked to a promoter element from P. pastoris and a Mxr1 transcriptional activator sequence from P. pastoris; and

a nucleic acid molecule encoding at least one polypeptide involved in heme biosynthesis operably linked to a promoter element from P. pastoris and a Mxr1 transcriptional activator sequence from P. pastoris.

2. The yeast cell of claim 1, wherein the heme-containing protein is Leghemoglobin (LegH).

3. The yeast cell of claim 2, wherein the LegH has the sequence shown in SEQ ID NO:6.

4. The yeast cell of claim 1, wherein the at least one polypeptide involved in heme biosynthesis is selected from the group consisting of ALA synthase, ALA dehydratase, porphogilinogen deaminase, UPG III synthase, UPG III decarboxylase, CPG oxidase, PPG oxidase, and/or ferrochelatase

5. The yeast cell of claim 1, wherein the promoter element from P. pastoris is a constitutive promoter element from P. pastoris or a methanol-inducible promoter element from P. pastoris.

6. The yeast cell of claim 5, wherein the constitutive promoter element from P. pastoris is a promoter element from a transcriptional elongation factor EF-1 (TEF1) gene.

7. The yeast cell of claim 5, wherein the methanol-inducible promoter element from P. pastoris is an alcohol oxidase I (AOX1) promoter element.

8. The yeast cell of claim 1, wherein the promoter element from P. pastoris comprises a constitutive promoter element from P. pastoris and a methanol-inducible promoter element from P. pastoris.

9. The yeast cell of claim 8, wherein the constitutive promoter element from P. pastoris is a promoter element from a transcriptional elongation factor EF-1 (TEF1) gene.

10. The yeast cell of claim 8, wherein the methanol-inducible promoter element from P. pastoris is an alcohol oxidase I (AOX1) promoter element.

11. The yeast cell of claim 1, wherein the nucleic acid molecule encoding the leghemoglobin is present in the yeast cell in at least two copies.

12. The yeast cell of claim 1, wherein the nucleic acid molecule encoding the leghemoglobin operably linked to the promoter element from P. pastoris and the Mxr1 transcriptional activator sequence from P. pastoris further comprises a termination sequence.

13. The yeast cell of claim 1, wherein the nucleic acid encoding the at least one polypeptide involved in heme biosynthesis operably linked to the promoter element from P. pastoris and the Mxr1 transcriptional activator sequence from P. pastoris further comprises a termination sequence.

ABSTRACT OF THE DISCLOSURE

Methods and materials for genetically engineering methylotrophic yeast are provided.

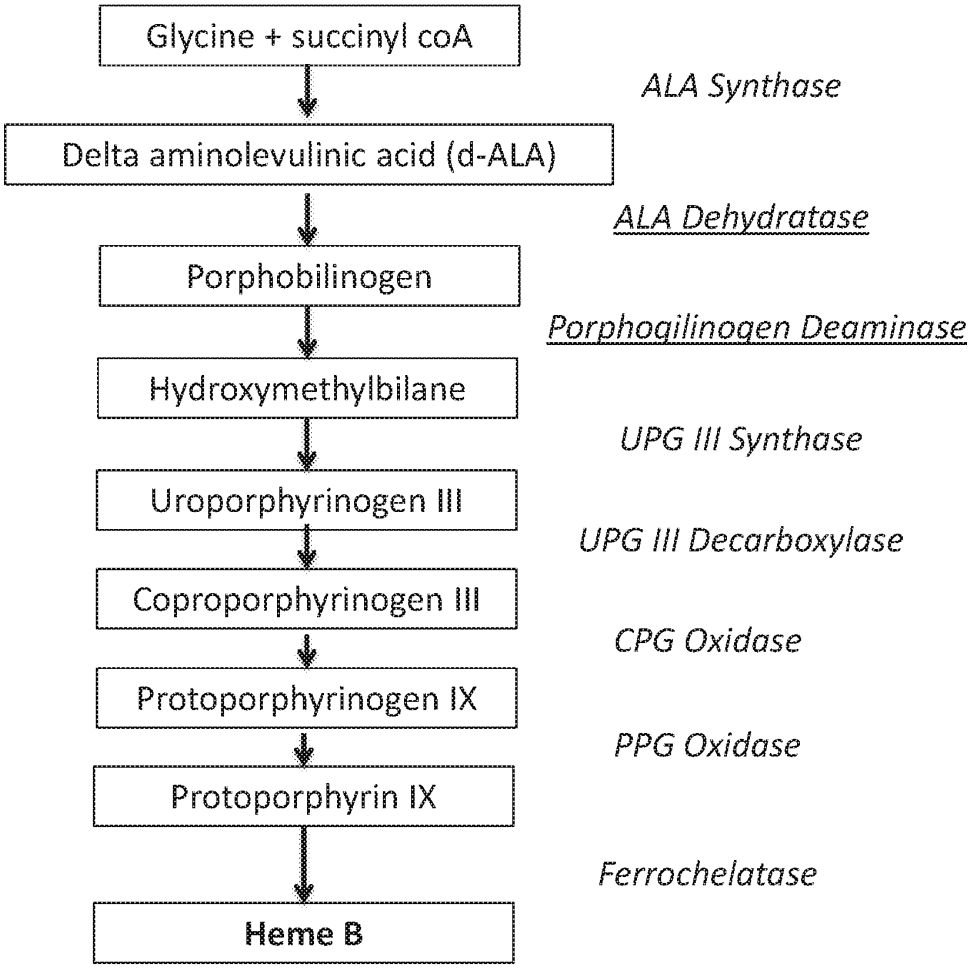

Glycine + succinyl coA
ALA Synthase
Delta aminolevulinic acid (d-ALA)
ALA Dehydratase
Porphobilinogen
Porphogilinogen Deaminase
Hydroxymethylbilane
UPG III Synthase
Uroporphyrinogen III
UPG III Decarboxylase
Coproporphyrinogen III
CPG Oxidase
Protoporphyrinogen IX
PPG Oxidase
Protoporphyrin IX
Ferrochelatase
Heme B

Figure 1

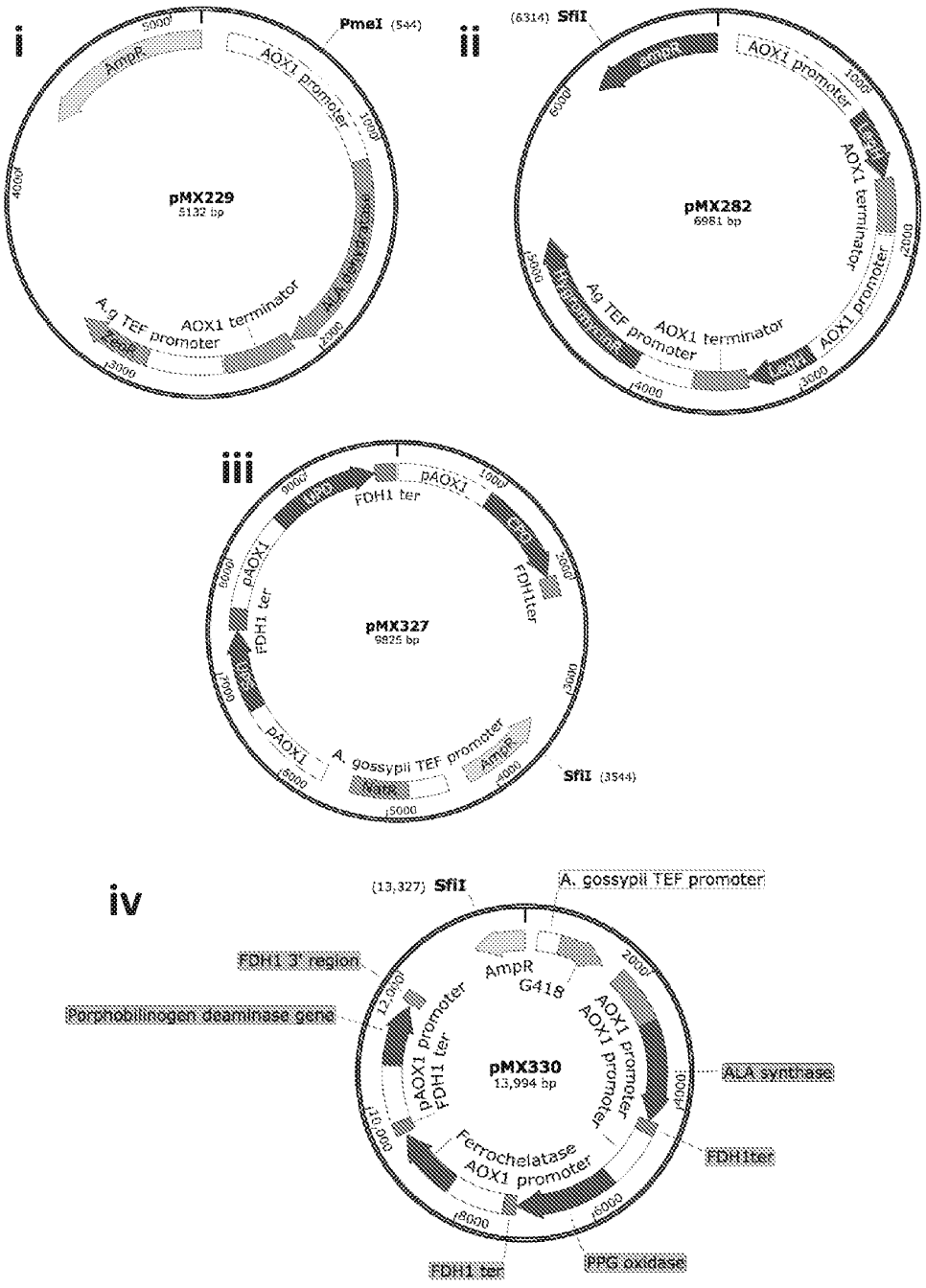

i
PmeI (544)
AmpR
AOX1 promoter
pMX229
5132 bp
A.g TEF promoter
AOX1 terminator
ii
(6314) SfiI
AmpR
AOX1 promoter
pMX282
6981 bp
A.g TEF promoter
AOX1 terminator
AOX1 promoter
iii
pAOX1
FDH1 ter
pMX327
9825 bp
A. gossypii TEF promoter
AmpR
SfiI (3544)
iv
(13,327) SfiI
A. gossypii TEF promoter
FDH1 3' region
Porphobilinogen deaminase gene
AmpR
G418
pMX330
13,994 bp
pAOX1 promoter
FDH1 ter
AOX1 promoter
ALA synthase
Ferrochelatase
FDH1ter
FDH1 ter
PPG oxidase

Figure 2

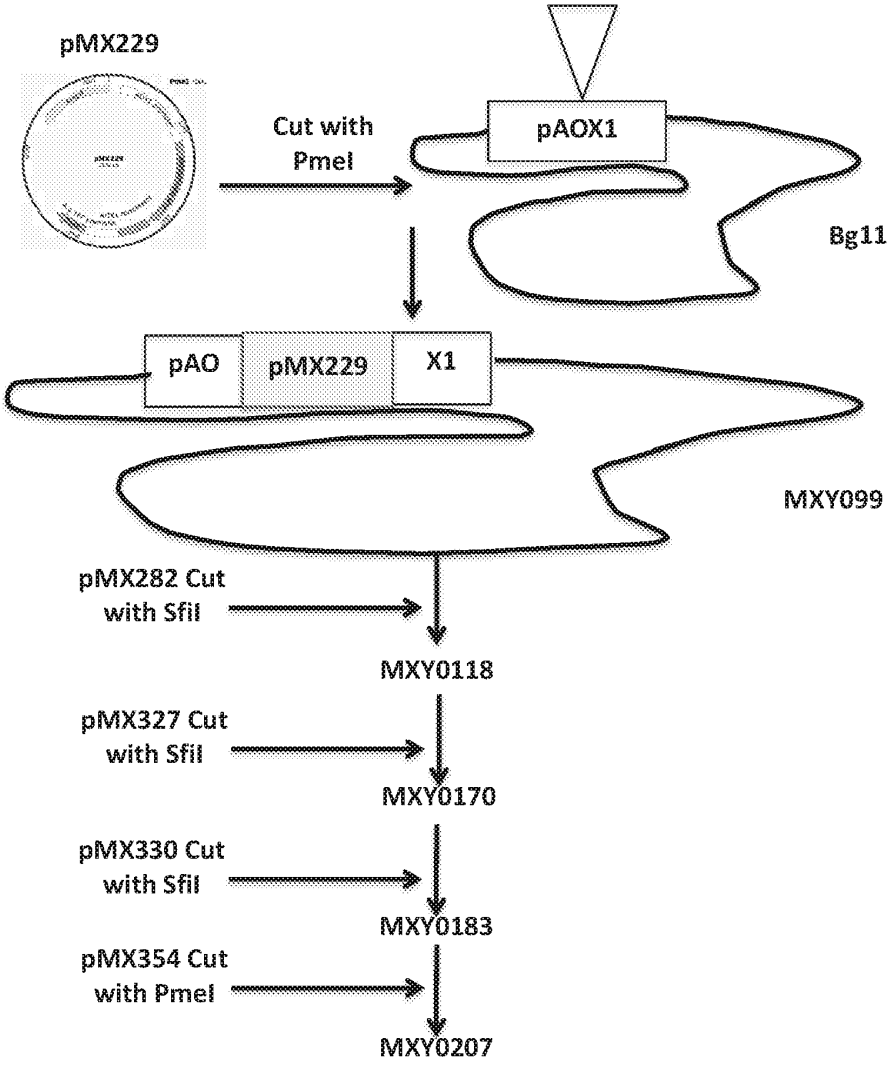

pMX229
Cut with
PmeI
pAOX1
Bg11
pAO
pMX229
X1
MXY099
pMX282 Cut
with SfiI
MXY0118
pMX327 Cut
with SfiI
MXY0170
pMX330 Cut
with SfiI
MXY0183
pMX354 Cut
with PmeI
MXY0207

Figure 3

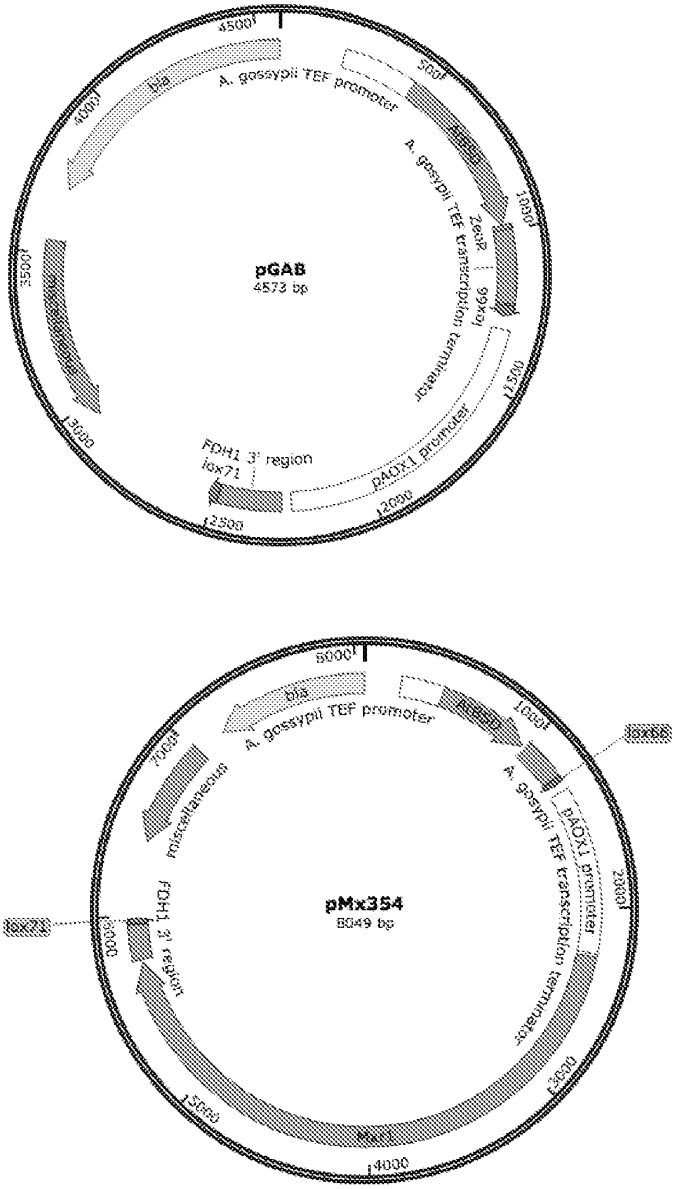

pGAB
4573 bp
bla
A. gossypii TEF promoter
A. gossypii TEF transcription terminator
ZeoR
loxP
pAOX1 promoter
FDH1 3' region
lox71
pMx354
8049 bp
bla
A. gossypii TEF promoter
ZeoR
A. gossypii TEF transcription terminator
pAOX1 promoter
FDH1 3' region
lox71
lox66
Kan/Neo
ori

Figure 4

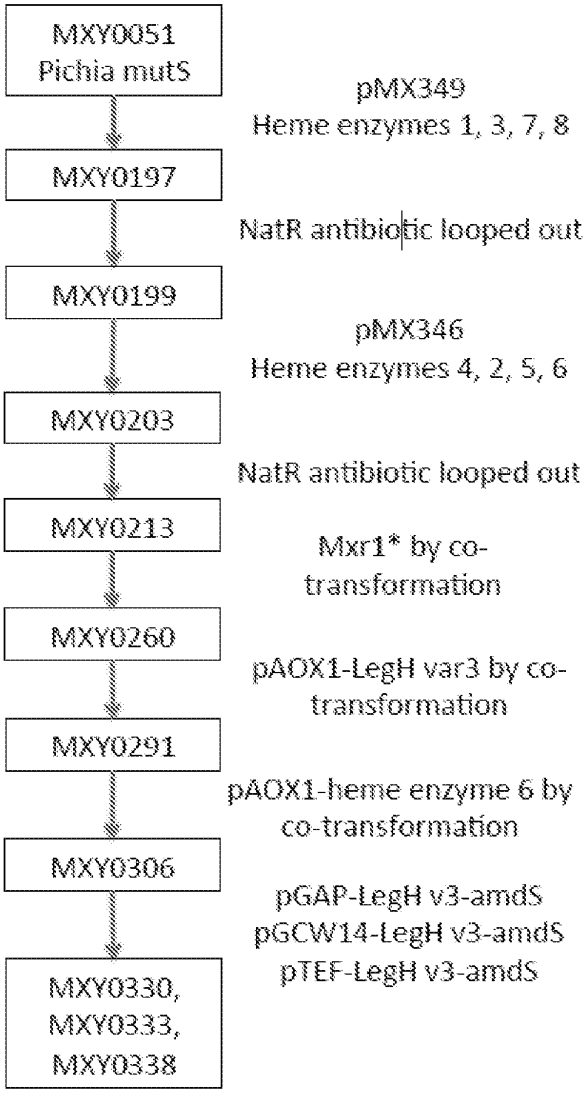

MXY0051
Pichia mutS
pMX349
Heme enzymes 1, 3, 7, 8
MXY0197
NatR antibiotic looped out
MXY0199
pMX346
Heme enzymes 4, 2, 5, 6
MXY0203
NatR antibiotic looped out
MXY0213
Mxr1* by co-transformation
MXY0260
pAOX1-LegH var3 by co-transformation
MXY0291
pAOX1-heme enzyme 6 by co-transformation
MXY0306
pGAP-LegH v3-amdS
pGCW14-LegH v3-amdS
pTEF-LegH v3-amdS
MXY0330,
MXY0333,
MXY0338

Figure 5

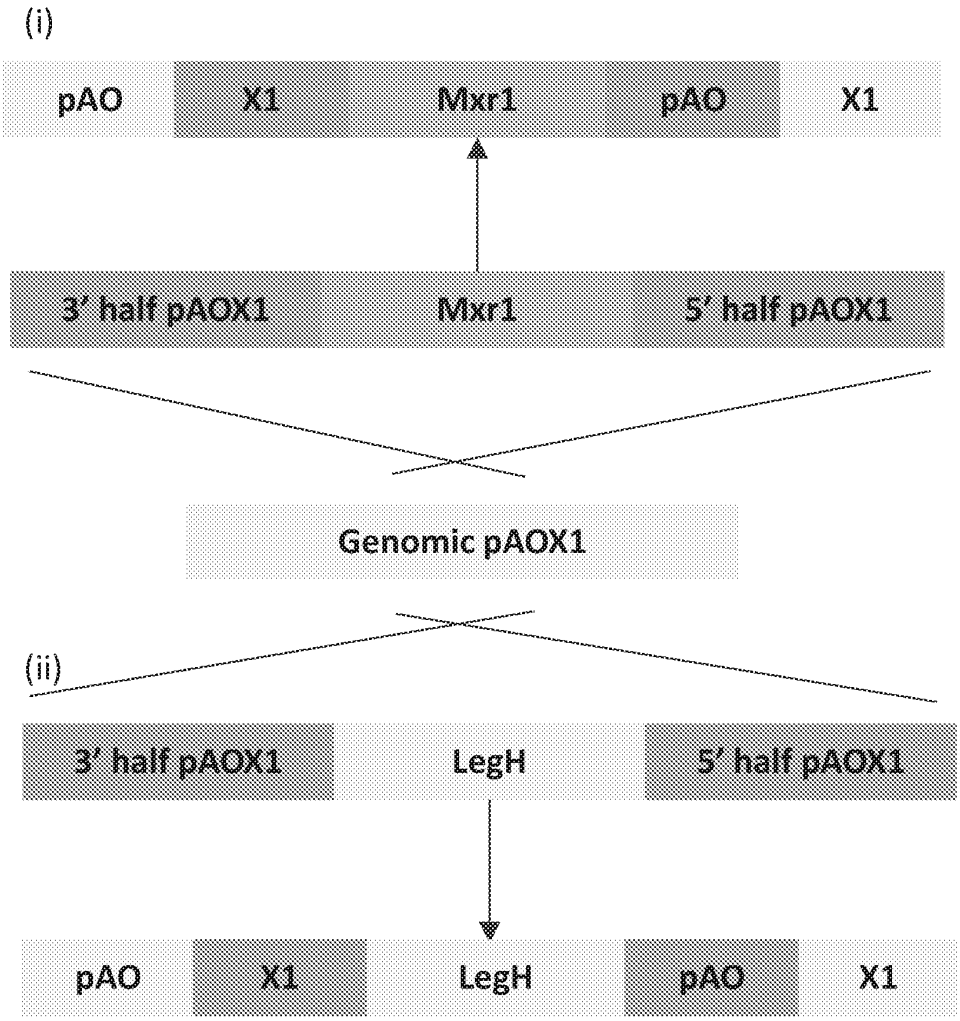

(i)
pAO
X1
Mxr1
pAO
X1
3′ half pAOX1
Mxr1
5′ half pAOX1
Genomic pAOX1
(ii)
3′ half pAOX1
LegH
5′ half pAOX1
pAO
X1
LegH
pAO
X1

Figure 6

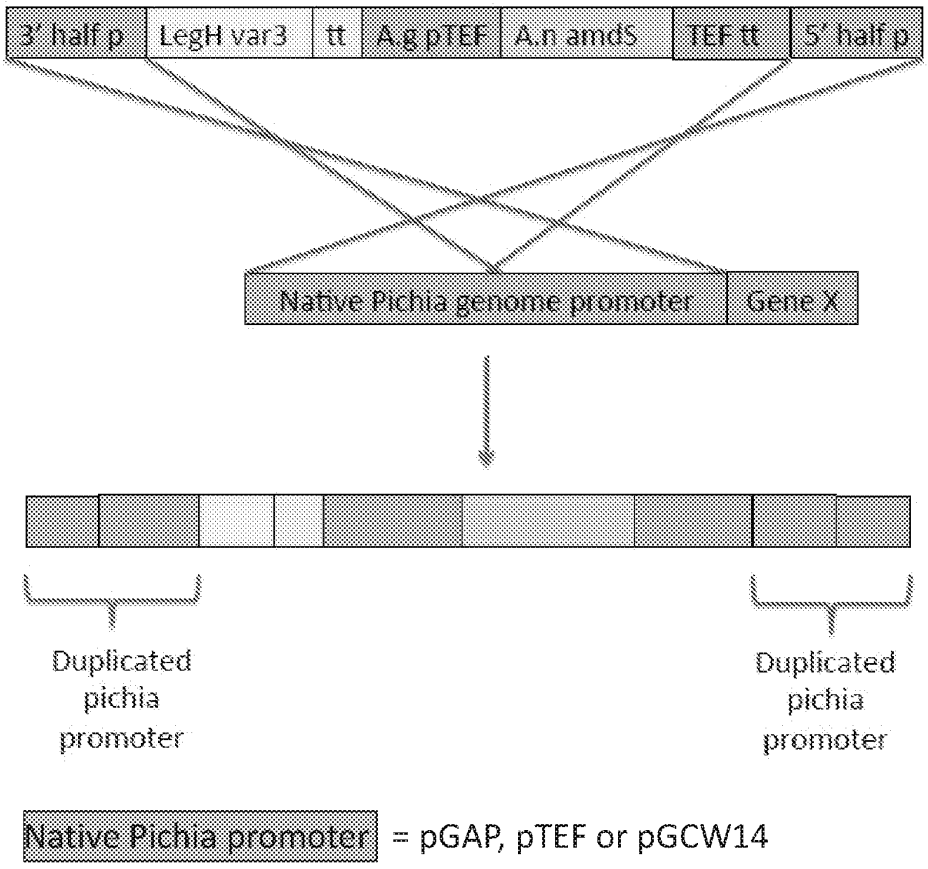

3' half p
LegH var3
tt
A.g pTEF
A.n amdS
TEF tt
5' half p
Native Pichia genome promoter
Gene X
Duplicated pichia promoter
Duplicated pichia promoter
Native Pichia promoter = pGAP, pTEF or pGCW14

Figure 7

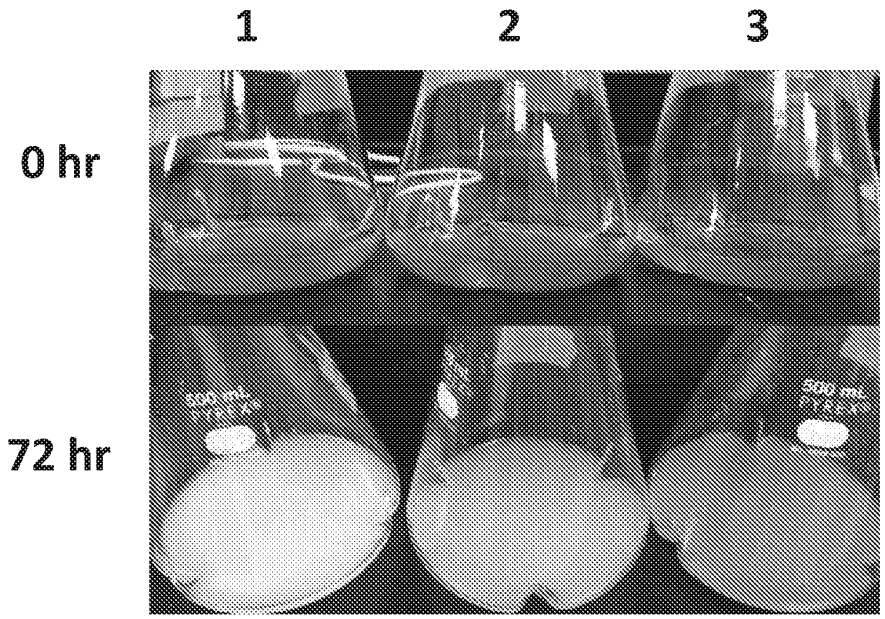
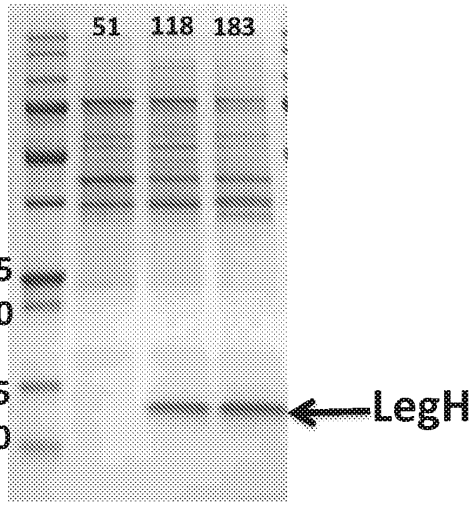


Figure 8

A

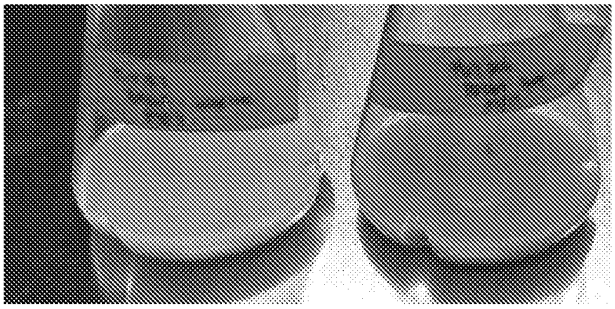


B

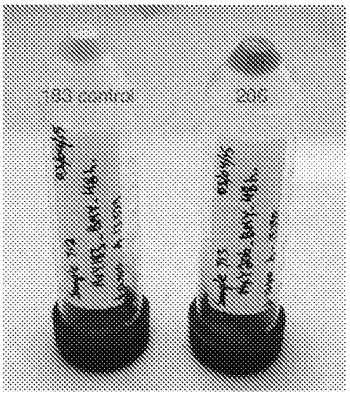
Strain	Relative improvement in yield (vs. MXY0118)	Heme-bound fraction of LegH
MXY0118	1	Sub-optimal
MXY0183	1.6	100%
MXY0207	3	100%

Figure 9

A



B



C

Strain	Relative yield of heme loaded LegH
183- BMGY	N/A
183- BMY	1
206- BMGY	3
206- BMY	5

Figure 10

MXY Strain	Yield on glycerol co-feed	Yield on dextrose co-feed
183	1X	1X
207	4.4X	4.4X
291	5.8X	5.8X
306	5.8X	5.8X
330	6.9X	6.9X
333	5.8X	5.8X
338	6.9X	6.9X

Figure 11

PTO/AIA/82A (07-13)
Approved for use through 11/30/2014. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5, unless the application number and filing date are identified in the Power of Attorney by Applicant form. If neither form PTO/AIA/82A nor form PTO/AIA82B identifies the application to which the Power of Attorney is directed, the Power of Attorney will not be recognized in the application.

Application Number	New
Filing Date	August 8, 2017
First Named Inventor	Smita Shankar and Martin Andrew Hoyt
Title	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST
Art Unit	
Examiner Name	
Attorney Docket Number	38767-0048002

SIGNATURE of Applicant or Patent Practitioner

Signature	/M. Angela Parsons/	Date (Optional)	8/16/2017
Name	M. Angela Parsons	Registration Number	44,282
Title (if Applicant is a juristic entity)			
Applicant Name (if Applicant is a juristic entity)			

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. If more than one applicant, use multiple forms.

☐ *Total of _____ forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**
61545145.doc

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

American LegalNet, Inc.
www.FormsWorkFlow.com

PTO/AIA/82B(07-13)
Approved for use through 11/30/2014. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

☒ I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above:

26191

OR

☐ I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter to:

☐ The address associated with the above-mentioned Customer Number

OR

☒ The address associated with Customer Number: **26191**

OR

Firm or Individual Name					
Address					
City		State		Zip	
Country					
Telephone		Email			

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

Impossible Foods Inc.

☐ Inventor or Joint Inventor (title not required below)

☐ Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)

☒ Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)

☐ Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature	/Pbr/	Date (Optional)	4/30/2014
Name	Patrick O'Reilly Brown M.D., Ph.D.		
Title	CEO, Impossible Foods Inc.		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

☐ *Total of _____ forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

American LegalNet, Inc.
www.FormsWorkFlow.com

PTO/AIA/82C (07-13)
Approved for use through 11/30/2014. OMB 0651-0051
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY BY APPLICANT

No more than ten (10) patent practitioners total may be appointed as set forth below by name and registration number. This page need not be submitted if appointing the Patent Practitioner(s) associated with a Customer Number (see form PTO/AIA/82B):

Name	Registration Number

American LegalNet, Inc.
www.FormsWorkFlow.com

DocCode – SEQ.TXT

SCORE Placeholder Sheet for IFW Content

Application Number: 15678342 Document Date: 08/16/2017

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Sequence Listing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

Attorney Docket No.: 38767-0048002

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor	:	Smita Shankar	Art Unit	:	Unknown
Serial No.	:	15/678,342	Examiner	:	Unknown
Filed	:	August 16, 2017			
Title	:	EXPRESSION CONSTRUCTS AND METHODS OF GENETICALLY ENGINEERING METHYLOTROPHIC YEAST			

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Please consider the references listed on the enclosed PTO-1449 Form. Foreign patent documents and non-patent literature are enclosed; U.S. patents and patent application publications will be provided upon request. Copies of Communications from a foreign Patent Office in counterpart applications also are enclosed.

This statement is being filed with the application. Therefore, no fees are believed due for this Information Disclosure Statement. If that is incorrect, please apply any charges, or credits, to Deposit Account No. 06-1050, referencing the above-indicated Attorney Docket Number.

Please contact the undersigned if there are any questions regarding this statement.

Respectfully submitted,

Date: /August 16, 2017/

/M. Angela Parsons/
M. Angela Parsons, Ph.D.
Reg. No. 44,282

Customer Number 26191
Fish & Richardson P.C.
Telephone: (612) 335-5070
Facsimile: (877) 769-7945

61545352.doc

Sheet 1 of 3

Substitute Disclosure Form	U.S. Department of Commerce Patent and Trademark Office	Attorney Docket No. 38767-0048002	Application No. 15/678,342
Information Disclosure Statement by Applicant (Use several sheets if necessary)		Applicant Impossible Foods Inc.	
(37 CFR §1.98(b))		Filing Date August 16, 2017	Group Art Unit

U.S. Patent Documents

Examiner Initial	Desig. ID	Document Number	Publication Date	Patentee	Class	Subclass	Filing Date If Appropriate
	1.	4,683,195	07/28/1987	Mullis et al.			
	2.	4,683,202	07/28/1987	Mullis			
	3.	4,800,159	01/24/1989	Mullis et al.			
	4.	4,965,188	10/23/1990	Mullis et al.			
	5.	5,204,253	04/20/1993	Sanford et al.			
	6.	5,538,800	07/23/1996	Jin et al.			
	7.	6,013,863	01/11/2000	Lundquist et al.			
	8.	6,329,571	12/11/2001	Hiei			
	9.	2002/00194643	12/19/2002	Bertrand et al.			
	10.	2008/10166757	06/01/2008	Bron et al.			

Foreign Patent Documents or Published Foreign Patent Applications

Examiner Initial	Desig. ID	Document Number	Publication Date	Country or Patent Office	Class	Subclass	Translation Yes	Translation No
	11.	WO 1998/012913	04/02/1998	WIPO				
	12.	WO 2004/057946	07/15/2004	WIPO				
	13.	WO 2004/099405	11/18/2004	WIPO				
	14.	WO 2012/083424	06/28/2012	WIPO				
	15.	WO 2013/010042	01/17/2013	WIPO				
	16.	WO 2014/110532	07/17/2014	WIPO				
	17.	WO 2014/110539	07/17/2014	WIPO				

Other Documents (include Author, Title, Date, and Place of Publication)

Examiner Initial	Desig. ID	Document
	18.	Chenna et al., "Multiple sequence alignment with the Clustal series of programs" Nucleic Acids Res., 31(13):3497-500 (2003)
	19.	Chiruvolu et al., "Recombinant protein production in an alcohol oxidase-defective strain of Pichia pastoris in fedbatch fermentations" Enzyme Microb. Technol., 21: 277-83 (1997)

Examiner Signature	Date Considered

EXAMINER: Initials citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

Substitute Disclosure Form

Substitute Disclosure Form U.S. Department of Commerce Patent and Trademark Office **Information Disclosure Statement by Applicant** (Use several sheets if necessary) (37 CFR §1.98(b))	Attorney Docket No. 38767-0048002	Application No. 15/678,342
	Applicant Impossible Foods Inc.	
	Filing Date August 16, 2017	Group Art Unit

Other Documents (include Author, Title, Date, and Place of Publication)		
Examiner Initial	Desig. ID	Document
	20.	Dayhoff et al., "A Model of Evolutionary Change in Proteins," Atlas of Protein Sequence and Structure, 5: 345-352 (1978)
	21.	GenBank Accession No. AB365355, "Candida boidinii TRM1 gene for Zn(II)2Cys6-type transcription factor Trm1, complete cds," dated January 24, 2008, 3 pages
	22.	GenBank Accession No. AB548760, "Candida boidinii TRM2 gene for C2H2-type transcription factor Trm2, complete cds," dated May 14, 2010, 3 pages
	23.	GenBank Accession No. ABD57365, "methanol expression regulator I [Komagataella pastoris]," dated March 4, 2006, 2 pages
	24.	GenBank Accession No. AEOI02000005, bases 858873 to 862352, "Ogataea parapolymorpha DL-1 chr3, whole genome shotgun sequence," dated December 3, 2013, 186 pages
	25.	GenBank Accession No. BAF99700, "Zn(II)2Cys6-type transcription factor Trm1 [Candida boidinii]," dated January 23, 2008, 2 pagges
	26.	GenBank Accession No. BAJ07608, "C2H2-type transcription factor Trm2 [Candida boidinii]," dated May 13, 2010, 2 pages
	27.	GenBank Accession No. DQ395124, "Pichia pastoris methanol expression regulator I gene, complete cds," dated March 4, 2006, 2 pages
	28.	GenBank Accession No. ESX01253, "Regulatory protein ADR1 [Ogataea parapolymorpha DL-1]," dated December 2, 2013, 2 pages
	29.	GenBank Accession No. FJ75255l, "Pichia pastoris dihydroxyacetone synthase 1 (DAS1) gene, complete cds" dated March 21, 2009, 2 pages
	30.	GenBank Accession No. X02425, "Hansenula polymorpha MOX gene for methanol oxidase" dated April 20, 1993, 3 pages
	31.	GenBank Accession No. YSAAODlA, "Candida boidinii methanol oxidase (AOD1) gene, complete cds," dated April 27, 1993, 2 pages
	32.	Hargrove et al., "Characterization of recombinant soybean leghemoglobin a and apolar distal histidine mutants" J Mol. Biol. 266, 1032-1042 (1997)
	33.	Inan & Meagher, "Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris" J. Biosci. Bioeng., 92:585-9 (2001)
	34.	International Search Report and Written Opinion in International Application No. PCT/US2014/055227, dated March 20, 2015, 20 pages
	35.	Jokipii-Lukkari et al., "Intrinsic non-symbiotic and truncated haemoglobins and heterologous Vitreoscilla haemoglobin expression in plants," Journal of Experimental Botany, 60(2):409-422 (January 2009)
	36.	Kranthi et al., "Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris" Yeast, 27:705-11 (2010)
	37.	Liachko & Dunham, "An autonomously replicating sequence for use in a wide range of budding yeasts" FEMS Yeast Res., 14:364-7 (2014)
	38.	Lloyd et al., "Targeted mutagenesis using zinc-finger nucleases in Arabidopsis" Proc. Natl. Acad. Sci. USA, 102:2232-2237 (2005)

Examiner Signature	Date Considered
EXAMINER: Initials citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.	

Substitute Disclosure Form

Sheet _3_ of _3_

Substitute Disclosure Form U.S. Department of Commerce Patent and Trademark Office	Attorney Docket No. 38767-0048002	Application No. 15/678,342
Information Disclosure Statement by Applicant (Use several sheets if necessary)	Applicant Impossible Foods Inc.	
(37 CFR §1.98(b))	Filing Date August 16, 2017	Group Art Unit

Other Documents (include Author, Title, Date, and Place of Publication)		
Examiner Initial	Desig. ID	Document
	39.	Londer et al. "Production and preliminary characterization of a recombiant triheme cytochrom c(7) from Geobatcher sulfurreducens in Escherichia coli." Biochem Biophy Act 1 1554(3): 202-211 (200)
	40.	Lutz et al., "A guide to choosing vectors for transformation of the plastid genome of higher plants" Plant Physiol. 145: 1201-1210 (2007)
	41.	Miele et al., "A GATA-type transcription factor regulates expression of the high-affinity iron uptake system in the methylotrophic yeast Pichia pastoris" Arch. Biochem. Biophys., 465: 172-9 (2007)
	42.	Nakagawa et al., "Alcohol oxidase hybrid oligomers formed in vivo and in vitro" Yeast, 15: 1223-30 (1999)
	43.	Qu et al., "Ectopic expression of the cottong non-symbiotic hemoglobin gene GhHb1 triggers defense responses and increases disease tolerance in Arabidopsis," Plant Cell Physiol., 47:1058-1068 (2006)
	44.	Raymond et al., "Development of the methylotrophic yeast Pichia methanolica for the expression of the 65 kilodalton isoform of human glutamate decarboxylase" Yeast, 14: 11-23 (1998)
	45.	Richards et al., "Construction of a GFP-BAR plasmid and its use for switchgrass transformation" Plant Cell. Rep. 20:48-20 54 (2001)
	46.	Sinagawa-Garcia et al., "Next generation synthetic vectors for transformation of the plastid genome of higher plants" Plant Afol Biol 70: 487-498 (2009)
	47.	Somleva et al., "Agrobacterium-Mediated Genetic Transformation of Switchgrass" Crop Sci. 42:2080-2087 (2002)
	48.	Supplementary European Search Report in International Application No. EP 14844701.4, dated January 27, 2017, 7 pages
	49.	Tanaka and Tanaka, "Tetrapyrrole biosynthesis in higher plants" Annu. Rev. Plant Biol. 58: 321-46 (2007)
	50.	Tovkach et al., "A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells" Plant J, 57: 747-757 25 (2009)
	51.	Townsend et al., "High-frequency modification of plant genes using engineered zinc-finger nucleases" Nature 459: 442-445 (2009)
	52.	UniprotKB Accession No. P02236, "Leghemoglobin C2," dated November 1, 1988, 9 pages
	53.	Wu et al., "Efficient production of a functional single-chain antidigoxin antibody via an engineered Bacillus subtilis expression-secretion system" Bio/Technology (NY) 11(1): 71-76 (1993) (abstract only)
	54.	Xie and Yang, "RNA-guided genome editing in plants using a CRISPR-Cas system" Mal. Plant 6: 1975-1983 (2013)
	55.	Zhang et al., "Transcription activator-like effector nucleases enable efficient plant genome engineering" Plant Physiology 161: 20-27 (2013)

Examiner Signature	Date Considered
EXAMINER: Initials citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.	

Substitute Disclosure Form

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :		(11) International Publication Number:
A01H 5/00, C12N 5/14, 15/31, 15/82	**A1**	**WO 98/12913**
		(43) International Publication Date: 2 April 1998 (02.04.98)

(21) International Application Number: PCT/US97/17246

(22) International Filing Date: 25 September 1997 (25.09.97)

(30) Priority Data:
08/720,260 26 September 1996 (26.09.96) US

(71)(72) Applicants and Inventors: BAILEY, James, E. [US/CH]; Winkelwiese 6, CH-8001 Zurich (CH). BULOW, Leif [SE/SE]; Borgaslingan 6, S-224 72 Lund (SE).

(74) Agents: ABRAMS, Samuel, B. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, GH, HU, ID, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: EXPRESSION OF GLOBIN PROTEINS IN PLANTS

(57) Abstract

The present invention relates to genetic-engineering of plants for enhanced oxygen assimilation and utilization. More particularly, this invention relates to producing transgenic plants engineered to express globin proteins such as, for example, hemoglobin, myoglobin, and hemoproteins. The engineered plants of the invention achieve quicker germination, are faster growing or maturing crops, produce higher crop yields, and/or contain higher levels of desired plant metabolites, particularly alkaloids. The invention also relates to mutant *Vitreoscilla* hemoglobin proteins, polynucleotides encoding the same, and host cells containing such polynucleotides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

EXPRESSION OF GLOBIN PROTEINS IN PLANTS

1. INTRODUCTION

The present invention relates to genetic-engineering of plants for enhanced oxygen assimilation and utilization. More particularly, this invention relates to producing transgenic plants engineered to express oxygen-binding proteins, particularly globins such as, for example, hemoglobin, myoglobin, and hemoproteins. The engineered plants may be used to achieve quicker germination, faster growing or maturing crops, higher crop yields and/or plants with higher levels of desired plant metabolites, particularly alkaloids.

2. BACKGROUND OF THE INVENTION

Genetic engineering of plants through transgenic approaches has rapidly become a diverse and promising field. Transformation techniques are applied both to study plant development and physiology, and in attempts to generate plants with advantageous agricultural properties.

For example, much effort has focused on improving crop production by transforming plants with herbicide or pesticide resistance genes. When grown under field conditions, the resulting transgenic plants can tolerate and survive applications of herbicides or pesticides which would destroy non-transformed plants. Alternatively, plants themselves have been genetically engineered to express products which are shown to act as natural insecticides such as *Bacillus thuringensis* toxin proteins (*see* U.S. Patent No. 5,380,831, issued Jan. 10, 1995).

Genetic engineering has also been attempted to generate plants with altered compositions of primary metabolites such as starch, sugar, and oils. For example, a bacterial levan sucrase gene was transformed into plants to increase levels of fructans in the resulting transformant. In another approach, the activity of certain enzymes involved in the metabolic synthesis of oils has been decreased using antisense technology in order to design vegetable oils exhibiting the desired properties. (*Reviewed in* Knauf, 1995, Curr. Opin. Biotech. 6:165-70).

In addition to attempts to alter primary metabolic compositions of plants (*i.e.* carbohydrates, oils, amino acids and proteins), it would also be desirable to alter secondary metabolite compositions. Secondary metabolites are those specialized plant products required by cells in small amounts (*e.g.* hormones), or

are highly specialized biomolecules (*e.g.* nucleotides, pigments, toxins, antibiotics, and alkaloids). Generally, secondary metabolites can be defined as compounds that have no recognized role in the maintenance of fundamental life processes of the cell, but which may perform other advantageous functions for the organism as a whole. (*See* for example, Bell, 1981, "The Physiological Role(s) of Secondary (Natural) Products" in The Biochemistry of Plants, v. 7, p. 1, Academic Press, Inc.)

The production of useful secondary metabolites from plants and plant cell cultures is an important aspect of plant technology. About 25% of the prescription drugs used in the industrialized world contain ingredients extracted from higher plants (Tyler, V.E., 1988, Planta Med. 54:95-100). The most prominent group of plant substances used in medicine are the alkaloids (Nickell, L.G., 1980, "Products," Plant tissue culture as a source of biochemicals, Staba, E.J., Ed., Boca Raton, CRC Press Inc., 235-269). Alkaloids are secondary metabolites which protect plants against phytophagocytosis by higher organisms or invasion by pathogens (Robins, R.J. *et al.*, 1991, Planta Med. 57:27-35). Plant alkaloids are generally derived from simple amino acids that interact with acetate and terpenoid units and undergo aromatic hydroxylations (Robins, *supra*). Examples of plant alkaloids are nicotine, scopolamine, hyoscyamine, ajmalicine, serpentine, piperine, leucenol, mimosine, ricinine, pelletierine, cocaine, hygrine, lupinine and anagyrine.

Several examples of introducing a new metabolic pathway or redirecting previously existing ones to increase production of secondary metabolites have been described (Bailey, *supra*; Berlin, J. *et al.*, 1994, Stud. Plant Sci. 4:57-81; Lilius, G., Holmberg, N. and Bülow, L., 1996, Bio/Technology 14:177-180). However, apparently all of these genetic engineering approaches entail either the up-regulation of a limiting enzyme in the pathway (for example through transformation and expression of the cloned gene), or the down-regulation of competing enzymatic pathways (for example through anti-sense technology). (*See* for review, Robins et al., 1991, Planta. Med. 57:27-35.)

Plant cell suspension cultures have often been used for industrial production of plant secondary metabolites. During *in vitro* culture, a critical parameter influencing secondary metabolite formation, especially in scale-up cultures, is the

level of dissolved oxygen. For instance, it has been demonstrated that a high dissolved oxygen concentration promotes the production of ajmalicine in *Catharantus roseus* batch cultures (Schaltmann, J.E. *et al.*, 1995, Biotech. Bioeng. 45:435-439). Similarly, increased alkaloid levels with increasing aeration have been observed in *Berberis wilsonae* (Breuling, M. *et al.*, 1985, Plant Cell Rep. 4:220-223). From these results it has been suggested that oxygen is a limiting substrate in plant tissue culture secondary metabolism, as well as for primary metabolism, especially for industrial scale-up processes. However, this need for increased oxygen has always been thought to be a need for extracellular oxygen during fermentation batch processes.

Aerobic metabolism has been successfully enhanced in fermentation microorganisms and cultured mammalian cells by engineering the host to express the *Vitreoscilla* hemoglobin gene (VHb) (Khosla, C., Bailey, J.E., 1988, Nature 331:633-635; Khosla, C., Bailey, J.E., 1988, Mol. Gen. Genet. 214:158-161; DeModena, *supra*; Chen, W. *et al.*, 1994, Biotechnol. Prog. 10:308-313; Pendse, G.J. and Bailey, J.E., 1994, Biotechnol. Bioeng. 44:1367-1370). This metabolic engineering strategy has been shown effective, for example, in increasing total cell protein synthesis by oxygen-limited *Escherichia coli* (Khosla, C. *et al.*, 1990, Bio/Technology 8:849-853 and U.S. Patent No. 5,049,493, issued Sept. 17, 1991), improving lysine yield and titer in cultivations of *Corynebacterium glutamicum* (Sander, F. *et al.*, 1994, Proc. 6th Eur. Congress Biotechnol., Alberghina, L., Frontali, L. and Sensi, P. (eds.) Elsevier Science B.V., Amsterdam 607-610), and in increasing actinorhodin and cephalosporin C production in *Streptomyces coelicolor* and *Acremonium chrysogenum*, respectively (Magnolo, S.K. *et al.*, 1991, Bio/Technology 9:473-476; DeModena, J.A. *et al.*, 1993, Bio/Technology 11:926-929).

Hemoglobin and globin-like proteins exist ubiquitously in mammals, and are less frequently found in plants. The few plant hemoglobin-like molecules described, which include lupin and soybean leghemoglobin, are largely thought to be associated with nitrogen-fixation activities of these plants, although some researchers claim that hemoglobin-like proteins occur in the roots of all plants. However, in no case have these proteins been overexpressed in plants. Although several major plant metabolic pathways are known to be oxygen-dependent,

including chlorophyll and heme biosynthesis, fatty acid desaturation and cysteine, glycine and serine biosynthesis, oxygen has not been thought to be a limiting factor in metabolism in intact plants. Rather, the observed effects of dissolved oxygen on metabolism in plant tissue cultures has been attributed to mass transport problems during large-scale, or even batch culture, fermentations (*see* Schlatmann et al., 1991, *supra*.). Thus, until the present invention, the advantages of altering useable levels of intracellular oxygen in intact plants have not been recognized.

3. SUMMARY OF THE INVENTION

The present invention is directed to the generation of plants with improved agronomic or medicinal characteristics by metabolically engineering an increased intracellular oxygen level or increased utilization of oxygen. Such engineered plants display enhanced production of valuable secondary metabolites, in particular those metabolites which are synthesized by oxygen-dependent pathways, as well as other unexpected and valuable properties such as more rapid germination, higher levels of chlorophyll, increased overall growth metabolism, and increased rate of accumulation of biomass.

The invention encompasses both genetically engineered plants and methods of producing the same. Therefore, in one of its embodiments, the invention encompasses a method of metabolically engineering plants by increasing intracellular levels of oxygen or intracellular utilization of oxygen. The invention includes a method of producing a plant with an improved agronomic or medicinal characteristic by transforming the plant with polynucleotides encoding an oxygen-binding protein. Particularly preferred oxygen-binding proteins are globins, which reversibly bind oxygen through a heme group. In an exemplary embodiment, the oxygen-binding protein is *Vitreoscilla* hemoglobin. Preferably, the polynucleotides encoding the oxygen-binding protein are operably linked to a strong, constitutively expressed plant promoter, for example, the CaMV 35S promoter. Another aspect of the invention is a plant or plant cell which has been genetically engineered by the methods of the invention. The invention encompasses plants and plant cells which are both dicotyledons and monocotyledons.

Still another aspect of the invention is mutant *Vitreoscilla* hemoglobin proteins, polynucleotides encoding the same, and host cells containing such mutant

proteins and polynucleotides. The host cells can be bacteria, plant, animal, fungi or algae. The invention also encompasses a method of generating polynucleotides encoding mutant *Vitreoscilla* hemoglobin proteins. The method entails amplifying a polynucleotide that encodes *Vitreoscilla* hemoglobin protein using a low fidelity polymerase chain reaction, transforming the amplified polynucleotides into bacteria, and selecting bacteria that contain a polynucleotide sequence encoding the mutant *Vitreoscilla* hemoglobin protein. In a preferred aspect of this method, bacterial colonies containing mutant *Vitreoscilla* hemoglobin proteins are selected by virtue of their color phenotype. Polynucleotides encoding mutant *Vitreoscilla* hemoglobin proteins that are isolated by this method, as well as the mutant proteins themselves, and host cells containing such polynucleotides, are also within the scope of the present invention.

4. BRIEF DESCRIPTION OF THE FIGURES

Figure 1: FIG. 1A. The last steps in the *Nicotiana tabaccum* biosynthesis of the secondary metabolites nicotine and anabasine. The reaction catalyzed by nicotine synthase, the oxygen-dependent enzyme catalyzing the conversion of nicotinic acid and N-methyl-Δ'-pyrrolinium salt into nicotine, is labelled in bold. FIG. 1B. Oxygen-dependent steps in chlorophyll and heme biosynthesis in plants. The last oxidation of chlorophyll a to chlorophyll b has only been observed *in vitro* (Scheer, H. (ed.), 1991, Chlorophylls CRC Press, Boca Raton, FL, USA 421-425).

Figure 2. Construction of the Ti-expression vector containing the *Vitreoscilla* VHb gene using PCR. The gene was placed under control of the CaMV 35S promoter and transcription termination was achieved by the NOS region. The approximate positions of the PCR primers are indicated by solid arrows. Right and left border of the T-DNA are indicated by RB and LB, respectively. NOS-prom, NPT II and NOS are abbreviations for the nopaline synthase promoter, the neomycine phosphotransferase II gene, and the nopaline synthase 3' transcription termination region, respectively.

Figure 3. Dried weight measurements at different time intervals are shown as mean values. Bars correspond to the standard error. The data from the first two time points (15 and 20 days) correspond to 12 plants but weighed in triplets,

and the data from the last two time points are from 8 plants individually weighed. This experiment was duplicated and the data points were subjected to statistical analysis using one-way analysis of variance (ANOVA).

Figure 4. Nicotine and anabasine production in the transgenic lines, He3 and He4, and a wild-type line. The nicotine concentrations are given μg/g fresh weight and the anabasine concentrations as μg/g freeze-dried plant. The nicotine and anabasine concentrations were measured in eight plants and six plants, respectively. All measurements were performed in duplicate, with the standard deviations being less than 1% and 25%, respectively.

Figure 5. The chlorophyll contents in the two transgenic tobacco lines and a wild-type control measured as μg chlorophyll/ml x mg fresh weight. The bars represent the mean values of 18 plants. The standard errors of the mean were less than 10%, 8% and 5% for wild-type tobacco, He3 and He4, respectively.

Figure 6. Schematic map of the pUC derivative plasmid pJFK2 used for transformation of rice.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to metabolic genetic engineering by effectively increasing the intracellular level of available oxygen in plants. The invention is based, in part, on the discovery that transforming plants so as to express oxygen-binding proteins, and in particular heme-containing globins, improves the properties of the transformed plants in unexpected ways. The resulting plants produce higher levels of secondary metabolites whose production is oxygen sensitive. Surprisingly, however, these genetically engineered plants also exhibit better growth characteristics, more rapid germination, and improved vegetative yield. Without intending to be limited to a particular mechanism, it is believed that these surprising results are due to either an increased level of intracellular oxygen, or an increased efficiency in intracellular oxygen utilization.

Accordingly, the targets for engineering are genes encoding proteins involved in binding oxygen, particularly those which bind oxygen reversibly such as the globins. The target genes include, but are not limited to, those encoding *Vitreoscilla* hemoglobin (VHb), horse heart myoglobin, E. coli hemoprotein, B.

subtilis hemoprotein, yeast flavohemoglobin, soybean leghemoglobin, lupin leghemoglobin, and sperm whale myoglobin. In fact, an engineered increased expression of any endogenous plant hemoglobin is also within the scope of the invention.

Particularly suitable for use in the present invention are those oxygen-binding proteins which have relatively high k_{off} rates such as VHb (k_{off} 5600 s^{-1}; Orii and Webster, 1986, J. Biol. Chem. 261:3544-3547) or relatively low oxygen affinity such as horse heart myoglobin (K_D 0.79 μM; Wittenberg et al., 1985, in Nitrogen fixation research progress, H.J. Evand et al. Eds. Martinus Nijhoff Publishers , Dordrecht, p. 354). Therefore, particularly preferred oxygen binding proteins may be those proteins with a k_{off} rate for oxygen of greater than 10 s^{-1}, or a K_D less than 0.5 μM, although it will be understood that oxygen-binding proteins with rate constants outside of these parameters will also be useful. Other examples of preferred oxygen-binding proteins are globins such as hemoglobin, myoglobin, and leghemoglobins. The properties of many oxygen-binding proteins, including globins, are disclosed in the literature. Additionally, techniques for determining the oxygen-binding properties of a protein such as a globin are well known to one of skill in the art and may be performed without undue experimentation.

A particularly advantageous oxygen-binding protein for use in the instant invention, as described herein by way of working example, is *Vitreoscilla* hemoglobin ("VHb"). The complete sequence of the VHb gene is described in U.S. Patent No. 5,049,493, *supra*. Mutants of VHb which bind oxygen are also within the scope of the present invention. As described more fully below by way of a working example, genes encoding functional mutants of VHb are generated through PCR mutagenesis, and then tested in a simple phenotypic screen for their ability to bind oxygen. Using these techniques, several VHb mutants have been obtained for use in the present invention.

Also encompassed within the scope of the invention are any nucleotide sequence which hybridizes to the complement of the nucleotide sequence of the VHb gene under highly stringent conditions, *e.g.*, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing

Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3) and encodes a functionally equivalent gene product; and any nucleotide sequence that hybridizes to the complement of the nucleotide sequence of the VHb gene under less stringent conditions, such as moderately stringent conditions, *e.g.*, washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*), yet which still encodes a functionally equivalent Vhb gene product. Functional equivalents of VHb include naturally occurring hemoglobin genes present in other species, and mutant VHb whether naturally occurring or engineered which retain at least some of the functional activity of VHb (*i.e.*, reversible oxygen binding). Species from which genes encoding functional equivalents of VHb include but are not limited to bacterial species, and in particular those bacterial species found in soil, dung, and in other oxygen-accessible, but poorly aerated, environments.

Furthermore, using the teachings of the present invention one of skill in the art can also isolate genes encoding other bacterial proteins which reversibly bind oxygen through a heme group without relying upon sequence homology with known proteins. For example, one may isolate and culture a bacterial organism, prepare genomic DNA from the cultured bacteria, and then construct an expression library with the DNA in *E. coli* using well known techniques. The expression library is plated to obtain isolated colonies. Those *E. coli* colonies carrying plasmids which expresses an oxygen-binding heme protein may be identified phenotypically by their red color. The expression library DNA clones encoding oxygen-binding proteins may be further characterized and subcloned into plant expression vectors. The expression of such oxygen-binding heme proteins in plants is accordingly encompassed by the invention.

The invention also encompasses degenerate variants of nucleotide sequences which encode the amino acid sequence of the VHb protein, VHb mutants, and functional equivalents of VHb encoded by nucleotide sequences which hybridize to the complement of the nucleotide sequence of the VHb gene. For example, the nucleotide sequence may be altered so as to optimize amino acid codon usage for expression in the chosen host cell.

The levels of oxygen-binding proteins can be altered or their expression can be enhanced or otherwise modified (*e.g.*, ectopic expression or tissue specific expression) to engineer a plant with desirable properties. The engineering is

accomplished by transforming plants with nucleic acid constructs described herein. The transformed plants or their progenies are screened for plants that express the desired protein or exhibit the desired altered expression of the oxygen-binding protein, altered expression of the corresponding mRNA, altered secondary metabolite levels, increased growth rate, enhanced vegetative yield and/or increased germination rate.

Engineered plants exhibiting the desired physiological and/or agronomic changes can be used in plant breeding or directly in agricultural production. These plants having one altered protein also may be crossed with other altered plants engineered with alterations in other pathways (*e.g.*, cross an oxygen-binding expressing plant to a plant which has been otherwise genetically engineered) to produce lines with even further enhanced physiological and/or agronomic properties compared to the parents.

The invention is illustrated by working examples of plants engineered for expression of *Vitreoscilla* hemoglobin (VHb). In all instances, engineered plants that exhibit expression also show better growth characteristics, increased rate of seed germination, improved vegetative yield and/or enhanced levels of secondary metabolites over control, wild-type plants.

The invention is based, in part, on the surprising discovery that transformation of plants with oxygen-binding proteins provides several advantages. Plants were transformed with expression plasmids encoding an oxygen-binding protein in an attempt to develop plants with increased drought tolerance. As a model system, *Nicotiana tabaccum* (tobacco) was first chosen to be transformed with expression constructs encoding the oxygen-binding protein *Vitreoscilla* hemoglobin (VHb). However, since there is an oxygen-dependent step in the nicotine biosynthesis pathway, the production of nicotine was examined.

The VHb gene was introduced into *Nicotiana tabaccum* (tobacco) via *Agrobacterium* mediated gene transfer. Transcription and translation of the VHb gene and message were demonstrated by reverse transcriptase PCR and Western immunoblot analysis, respectively. Transgenic tobacco plants expressing VHb exhibited on average 34% greater nicotine content than wild-type controls. They also contained an altered distribution of nicotine relative to the less desirable alkaloid anabasine. Additional effects of VHb expression on the physiology of the

transgenic tobacco plants were also noted; these plants contained on average 10-20% more chlorophyll than controls. Moreover, the plants germinated more quickly and exhibited a more rapid rate of dry weight increase. Thus, engineering of plants to express additional oxygen-binding proteins not only produced increased alkaloid levels and altered distributions of alkaloids, but also enhanced plant growth.

In yet another illustration of the advantages of the present invention, plants from the genus *Datura* were transformed with the same expression construct used to transform the tobacco plants. *Datura* is the major commercial source for production of the secondary metabolites scopolamine and hyoscyamine, which are used as sedatives and as starting material for semisynthetic derivatives of these metabolites. Surprisingly, scopolamine and hyoscyamine levels in the transgenic plants were increased as much as 5 to 6 fold over levels found in non-transformed plants.

In still another illustration of the invention, plants from the genus *Oryzae* were transformed by the ballistic method with a polynucleotide construct that directs the expression of the VHb hemoglobin gene. Thus, the present invention is generally applicable for altering the oxygen utilization in all plant types in order to obtain plant varieties with advantageous medicinal and agronomic qualities. For example, plants which germinate more quickly and accumulate biomass more rapidly will find particular application in agricultural environments where arable land and/or the length of growing season are limiting factors to production. Even a small increase in biomass production and/or rate of growth can prove significant. Additionally, the increased levels of economically important secondary metabolites may enable production of metabolites from previously economically unfeasible plant sources.

Further, the invention encompasses a method of generating polynucleotides encoding mutant *Vitreoscilla* hemoglobin proteins. The method entails amplifying a polynucleotide that encodes *Vitreoscilla* hemoglobin protein using a low fidelity polymerase chain reaction, transforming the amplified polynucleotides into bacteria, and selecting bacteria that contain a polynucleotide sequence encoding the mutant *Vitreoscilla* hemoglobin protein. In a preferred aspect of this method, bacterial colonies containing mutant *Vitreoscilla* hemoglobin proteins are selected

by virtue of their color phenotype; a red or pink color indicates that the bacteria express an oxygen-binding heme protein. Polynucleotides encoding mutant *Vitreoscilla* hemoglobin proteins that are isolated by this method, as well as the mutant proteins themselves, and host cells containing such polynucleotides, are within the scope of the present invention. The host cells can be bacteria, plant, animal, fungi or algae. Such mutant *Vitreoscilla* hemoglobin proteins are useful in any number of applications for which wild-type *Vitreoscilla* hemoglobin proteins are advantageous, including but not limited to increasing cell growth and increasing cell production.

5.1. PRODUCTION OF TRANSGENIC PLANTS

According to the present invention, a desirable plant or plant cell may be obtained by transforming a plant cell with the nucleic acid encoding an oxygen-binding protein, particularly a globin, and in one embodiment VHb, as described herein. In some instances, it may be desirable to engineer a plant with several different gene constructs. Such engineering may be accomplished by transforming a plant or plant cell with all of the desired gene constructs simultaneously. Alternatively, the engineering may be carried out sequentially. That is, transforming with one gene construct, obtaining the desired transformant after selection and screening, transforming the transformant with a second gene construct, and so on.

In an embodiment of the present invention, *Agrobacterium* is employed to introduce the gene constructs into plants. Such transformations preferably use binary *Agrobacterium* T-DNA vectors (Bevan, 1984, Nuc. Acid Res. 12:8711-8721), and the co-cultivation procedure (Horsch et al., 1985, Science 227:1229-1231). Generally, the *Agrobacterium* transformation system is used to engineer dicotyledonous plants (Bevan et al., 1982, Ann. Rev. Genet. 16:357-384; Rogers et al., 1986, Methods Enzymol. 118:627-641) The *Agrobacterium* transformation system may also be used to transform, as well as transfer, DNA to monocotyledonous plants and plant cells (*see* Hernalsteen et al., 1984, EMBO J. 3:3039-3041; Hooykass-Van Slogteren et al., 1984, Nature 311:763-764; Grimsley et al., 1987, Nature 325:1677-179; Boulton et al., 1989, Plant Mol. Biol. 12:31-40; Gould et al., 1991, Plant Physiol. 95:426-434).

In other embodiments, various alternative methods for introducing recombinant nucleic acid constructs into plants and plant cells may also be utilized. These other methods are particularly useful where the target is a monocotyledonous plant. Alternative gene transfer and transformation methods include, but are not limited to, protoplast transformation through calcium-, polyethylene glycol (PEG)- or electroporation-mediated uptake of naked DNA (*see* Paszkowski et al., 1984, EMBO J 3:2717-2722, Potrykus et al., 1985, Mol. Gen. Genet. 199:169-177; Fromm et al., 1985, Proc. Natl. Acad. Sci. USA 82:5824-5828; Shimamoto, 1989, Nature 338:274-276) and electroporation of plant tissues (D'Halluin et al., 1992, Plant Cell 4:1495-1505). Additional methods for plant cell transformation include microinjection, silicon carbide mediated DNA uptake (Kaeppler et al., 1990, Plant Cell Reporter 9:415-418), and microprojectile bombardment (*see* Klein et al., 1988, Proc. Natl. Acad. Sci. USA 85:4305-4309; Gordon-Kamm et al., 1990, Plant Cell 2:603-618).

According to the present invention, a wide variety of plants may be engineered for the desired physiological and agronomic characteristics described herein using nucleic acid constructs encoding oxygen-binding proteins and the various transformation methods mentioned above. Target plants for engineering include, but are not limited to, those of tobacco, *Datura*, maize, wheat, rice, soybean, tomato, carrots, peanut, potato, sugar beets, sunflower, yam, *Arabidopsis*, rape seed, petunia and other ornamental flowers, turfgrass, barley, alfalfa, potato, cassava, sweet potato, lettuce, sugar cane, fruit trees (*e.g.* cherry, apple, peach, pear, plum, orange, lemon), nut trees (*e.g.* walnut, almond), banana, peanut, strawberry, cauliflower, artichoke, cucumber, melon, grape, pepper, cabbage, endive, leek, lettuce, spinach, arrowroot, radish, bean and pea.

According to the present invention, desired plants may be obtained by engineering the gene constructs described herein into a variety of plant cell types, including but not limited to, protoplasts, tissue culture cells, tissue and organ explants, pollen, embryos as well as whole plants. In an embodiment of the present invention, the engineered plant material is selected or screened for transformants (*i.e.*, those that have incorporated or integrated the introduced gene construct(s)) following the approaches and methods described below. An isolated transformant may then be regenerated into a plant. Alternatively, the engineered

plant material may be regenerated into a plant, or plantlet, before subjecting the derived plant, or plantlet, to selection or screening for the marker gene traits. Procedures for regenerating plants from plant cells, tissues or organs, either before or after selecting or screening for marker gene(s), are well known to those skilled in the art.

A transformed plant cell, callus, tissue or plant may be identified and isolated by selecting or screening the engineered plant material for traits encoded by the marker genes present on the transforming DNA. For instance, selection may be performed by growing the engineered plant material on media containing inhibitory amounts of the antibiotic or herbicide to which the transforming marker gene construct confers resistance. Further, transformed plants and plant cells may also be identified by screening for the activities of any visible marker genes (*e.g.*, the ß-glucuronidase, luciferase, B or C1 genes) that may be present on the recombinant nucleic acid constructs of the present invention. Such selection and screening methodologies are well known to those skilled in the art.

Physical and biochemical methods may also be used to identify a plant or plant cell transformant containing the gene constructs of the present invention. These methods include but are not limited to: 1) Southern analysis or PCR amplification for detecting and determining the structure of the recombinant DNA insert; 2) northern blot, S-1 RNase protection, primer-extension or reverse transcriptase-PCR amplification for detecting and examining RNA transcripts of the gene constructs; 3) enzymatic assays for detecting enzyme or ribozyme activity, where such gene products are encoded by the gene construct; 4) protein gel electrophoresis, western blot techniques, immunoprecipitation, or enzyme-linked immunoassays, where the gene construct products are proteins; 5) biochemical measurements of compounds produced as a consequence of the expression of the introduced gene constructs. Additional techniques, such as *in situ* hybridization, enzyme staining, and immunostaining, may also be used to detect the presence or expression of the recombinant construct in specific plant organs and tissues. The methods for doing all these assays are well known to those skilled in the art.

5.1.1. PLANTS TRANSFORMED TO EXPRESS OXYGEN-BINDING PROTEINS

In accordance with the present invention, a plant that expresses a recombinant oxygen-binding protein gene may be engineered by transforming a plant cell with a gene construct comprising a plant promoter operably associated with a sequence encoding oxygen-binding protein. (Operably associated is used herein to mean that transcription controlled by the "associated" promoter would produce a functional messenger RNA, whose translation would produce the enzyme.) The plant promoter may be constitutive or inducible. Useful constitutive promoters include, but are not limited to, the CaMV 35S promoter, the T-DNA mannopine synthetase promoter, and their various derivatives. Useful inducible promoters include but are not limited to the promoters of ribulose bisphosphate carboxylase (RUBISCO) genes, chlorophyll a/b binding protein (CAB) genes, heat shock genes, the defense responsive gene (*e.g.*, phenylalanine ammonia lyase genes), wound induced genes (*e.g.*, hydroxyproline rich cell wall protein genes), chemically-inducible genes (*e.g.*, nitrate reductase genes, gluconase genes, chitinase genes, PR-1 genes *etc.*), dark-inducible genes (*e.g.*, asparagine synthetase gene (Coruzzi and Tsai, U.S. Patent 5,256,558, October 26, 1993, Gene Encoding Plant Asparagine Synthetase) to name just a few.

In yet another embodiment of the present invention, it may be advantageous to transform a plant with a gene construct operably linking a modified or artificial promoter to a sequence encoding oxygen-binding protein. Typically, such promoters, constructed by recombining structural elements of different promoters, have unique expression patterns and/or levels not found in natural promoters. *See, e.g.* Salina et al., 1992, Plant Cell 4:1485-1493, for examples of artificial promoters constructed from combining cis-regulatory elements with a promoter core.

In a preferred embodiment of the present invention, the associated promoter is a strong and/or embryo-specific plant promoter such that the oxygen-binding protein is overexpressed in the transgenic plant. Additionally, particular promoters may be used to direct tissue specific expression of the oxygen-binding protein. For example, root-specific expression of an oxygen-binding protein may be engineered.

Further, recombinant expression of an oxygen-binding protein may be directed to specific cellular compartments. One of skill in the art may direct expression of the oxygen-binding protein intracellularly, to the chloroplasts, or to the mitochondria, to give just a few examples.

In yet another preferred embodiment of the present invention, the overexpression of oxygen-binding protein in plants may be engineered by increasing the copy number of the oxygen-binding protein gene. One approach to producing such transgenic plants is to transform with nucleic acid constructs that contain multiple copies of a plant expression construct directing the expression of an oxygen-binding protein gene (*i.e.*, operatively linked to a plant promoter). Another approach is to repeatedly transform successive generations of a plant line with one or more copies of the plant expression construct. Yet another approach is to place an oxygen-binding protein expression construct in a nucleic acid construct containing an amplification-selectable marker (ASM) gene such as the glutamine synthetase or dihydrofolate reductase gene. Cells transformed with such constructs are subjected to culturing regimes that select cell lines with increased copies of the oxygen-binding protein gene. *See* Donn et al., 1984, J. Mol. Appl. Genet. 2:549-562, for a selection protocol used to isolate of a plant cell line containing amplified copies of the GS gene. Because the desired gene is closely linked to the ASM, cell lines that amplified the ASM gene are also likely to have amplified the desired oxygen-binding protein gene. Cell lines with amplified copies of the oxygen-binding protein gene can then be regenerated into transgenic plants.

5.1.2. SCREENING OF TRANSFORMED PLANTS FOR THOSE HAVING DESIRED ALTERED TRAITS

It will be recognized by those skilled in the art that in order to obtain transgenic plants having the desired engineered traits, screening of transformed plants (i.e., those having an gene construct of the invention) having those traits may be required. For example, where the plants have been engineered for overexpression of an oxygen-binding protein gene, transformed plants are examined for those expressing the oxygen-binding protein gene at the desired level and in the desired tissues and developmental stages. The plants exhibiting the

desired physiological changes, *e.g.*, oxygen-binding protein expression, may then be subsequently screened for those plants that have the desired phenotypic changes at the plant level (e.g., increased chlorophyll, heme, amino acid, or alkaloid biosynthesis, increased fatty acid desaturation, or increased levels of oxmoprotectants such as glycine betaine which requires oxygen for its biosynthesis). The same principle applies to obtaining transgenic plants having tissue-specific expression of a heterologous gene by the use of a tissue specific promoter driven expression construct.

Alternatively, the transformed plants may be directly screened for those exhibiting the desired structural changes. In one embodiment, such screening may be for the color of the transformed plants. In other embodiments, the screening of the transformed plants may be for improved agronomic characteristics (e.g., faster growth, greater vegetative or reproductive yields, or improved protein contents, etc.), as compared to unengineered progenitor plants, when cultivated under various growth conditions (e.g., soils or media containing different amount of nutrients, water content, and/or salt content).

According to the present invention, plants engineered with oxygen-binding protein expression constructs, particularly *Vitreoscilla* hemoglobin, may exhibit improved vigorous growth characteristics, increased levels of chlorophyll, and of secondary plant metabolites, particularly alkaloids.

Engineered plants and plant lines possessing such improved agronomic characteristics may be identified by examining any of following parameters: 1) the rate of growth, measured in terms of rate of increase in fresh or dry weight, or rate of increase in height; 2) vegetative yield of the mature plant, in terms of fresh or dry weight; 3) the rate of germination; 4) the level of chlorophyll content; 5) the level of a secondary metabolite of interest, such as nicotine, scopolamine, hyoscyamine; and 6) the color of the plant. The procedures and methods for examining these parameters are well known to those skilled in the art.

According to the present invention, a desired plant is one that exhibits improvement over the control plant (*i.e.*, progenitor plant) in one or more of the aforementioned parameters. In an embodiment, a desired plant is one that shows at least 5% increase over the control plant in at least one parameter. In a preferred embodiment, a desired plant is one that shows at least 20% increase over the

control plant in at least one parameter. Most preferred is a plant that shows at least 50% increase in at least one parameter.

6. EXAMPLE: EXPRESSION OF VITREOSCILLA HEMOGLOBIN IN TOBACCO

An extensively investigated plant alkaloid is nicotine, the principal alkaloid found in tobacco (*Nicotiana tabaccum*). The biosynthetic pathways for the tobacco alkaloids have become accepted but regulation of this pathway is not yet fully understood, and all of its enzymes are not characterized (Bush, L.P., 1981, Recent Adv. Tob. Sci. 7:75-106). However, the activity of the final enzyme in nicotine biosynthesis, nicotine synthase, which catalyzes the formation of (s)-nicotine from nicotinic acid and N-methyl-Δ'-pyrrolinium chloride, has proven to be oxygen-dependent (Figure 1) (Friesen, J.B. and Leete, E., 1990, Tetrahedron Lett. 31:6295-6298), and the competing pathway from nicotinic acid and N-methyl-Δ'-pyrrolinium chloride to anabasine is not, so far as is known, stimulated by oxygen. Thus, this system was chosen as a first experiment to investigate the effects of increasing oxygen availability by providing oxygen-binding proteins to the plant. In this example, tobacco was transformed with a VHb expression vector via *Agrobacterium* mediated gene transfer. Subsequent assays were performed to determine: (i) if the VHb gene was correctly transcribed and translated in tobacco, (ii) the response of nicotine and anabasine accumulation to VHb expression, and, based on other features observed for the VHb-expressing transgenic plants, (iii) chlorophyll content and growth rate of VHb containing plants.

6.1. Experimental protocol

Recombinant DNA techniques and bacterial strains

Plasmid procedures were performed according to the methods described in Sambrook, J. *et al.*, 1990, Molecular cloning. A laboratory manual, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press, and PCR was performed using standard protocols (Saiki, *supra*). *Escherichia coli* strain TGI (F'*traD36lacI9Δ[lacZ]M15proA+B+/supE[hsdM-mcrB]5[r_K-m_K-McrB-]thi[lac-proAB][lac-prof*) was cultivated and transformed, using standard techniques (Sambrook, J., *supra*). Restriction enzymes were used according to the suppliers'

recommendations (GIBCO BRL, Labdesign). *Agrobacterium tumefaciens* strain LBA 4404 (Walden, R. *et al.*, 1990, MMCB 1:175-194) was cultivated in LB medium (tryptone 10 g/l, yeast extract 5g/l, NaCl 10 g/l) and transformed with the aid of *E.coli* HB 101, harboring conjugative plasmid pRK 2013 (Figurski, D.H. and Helinski, D.R., 1979, Proc. Natl. Acad. Sci. U.S.A. 76:1648-1652). Extraction of genomic tobacco DNA for PCR analysis was performed as described by Edwards, K. *et al.*, 1991, Nucleic Acids Res. 19:1349).

DNA-constructions

PCR was used to amplify the VHb gene from plasmid pINT1 (Khosla, C. and Bailey J.E., 1989, J. Mol. Biol. 210:79-90). The 5' primer H1 (with sequence 5'- GCA AAC AGG ATC CCG CGT CTA GAG G 3') was complementary to a region upstream of the transcription initiation codon containing several restriction enzyme sites, the 3' primer H2 (with sequence 5'- GCA GAT TTG TAC GCT CAA GCG GTT GAA TGA GCT CCC AAA G - 3') was complementary to the coding region of the VHb gene upstream of the stop codon; primers were synthesized at the Biomolecular Unit, Lund University. The stop codon was also altered, using PCR mediated mutagenesis, from TAA to TGA, which is a more frequently used stop codon in tobacco. The amplified PCR fragment was inserted into pUC19 which was subsequently transformed into *E. coli*. Bacteria expressing the VHb gene exhibit a red tint when pelleted. One of the best expressing *E. coli* clones, pUH4, was chosen for further experiments. To verify that the cloned PCR fragment of the VHb gene expressed a functional protein, difference spectrum analysis was used (Webster, D.A. and Liu, C.Y., 1974, J. Biol. Chem. 249:4257-4260). Analysis of the CO spectrum of recombinant *E. coli* harboring pUH4 revealed a pronounced peak at 419 nm which is characteristic for the VHb activity (Lilius, *supra*). The VHb gene fragment was cleaved out with *XbaI* and *SacI* and ligated into plasmid pBin121:2 (Jefferson, R.A., 1987, Plant. Mol. Biol. Rep. 5:387-405) giving pBH4 (Figure 2).

Plant material

Nicotiana tabaccum SR. plants were exposed to 16 h light at 25°C and were either kept in soil or in sterile jars on MS-media (Murashige, T. and Skoog, F., 1962, Physiol. Plant 15:473-497). The tobacco plants were transformed by *Agrobacterium*-mediated gene transfer as described by Rogers, S.G. and Fraley,

R.T., 1986, Methods Enzymol. 118:627-684. Stable heterozygous transformants of the F_1 generation were used in all plant experiments.

Transcription analysis using RT-PCR

Polyadenylated mRNA was isolated from 100 mg leaf material using oligo(dT)-cellulose as in the Pharmacia Quick Prep™ manual. The VHb gene transcript was amplified by *rTth DNA polymerase* and the primers according to the instructions given by Perkin Elmer.

Western immunoblot analysis

Protein extracts of two transgenic tobacco lines and a wild-type control were prepared as follows: One gram leaf and stem material was homogenized in 0.1 M sodium phosphate, 5 mM DTT and 5 mM EDTA buffer at pH 7.5. The homogenates were centrifuged in 1.5 mL test tubes at 12,000 rpm for 15 min; the protein-containing supernatants were decanted and used for further studies. The proteins were separated on 15 % SDS-PAGE gels and blotted onto Immobilon P nylon membranes supplied by Millipore. The electrophoreses and blotting were performed as described by Laemmli (Laemmli, U.K., 1970, Nature 277:680-685) and Sambrook, *supra*, respectively. The VHb monomers were detected by VHb antiserum generated in rabbit, and visualized by peroxidase conjugated swine anti-rabbit immunoglobulins (Dako).

Alkaloid analysis

Transgenic tobacco and wild-type plants were analyzed for their nicotine and anabasine levels. The plants were grown for 30 days on MS medium, then harvested and homogenized in approximately 40 mL 0.1 M Tris-HCl buffer pH 9.0, centrifuged and stored at -70°C. The nicotine and anabasine concentrations were determined using capillary column gas chromatography as described by Curvall *et al.*, 1982, J. Chrom. 232:283-293.

Chlorophyll assay

Leaf material from three-week old seedlings was homogenized in 1.5 ml tubes and the chlorophyll was extracted with ice-cold 80% acetone. The debris was removed and the supernatants were collected. The absorbance of the chlorophyll extracts were monitored at 645 and 633 nm. The absorbencies were normalized to the amount of plant material used in each extraction. The amounts

of chlorophyll-A and B were calculated as described by Arnon, D.I., 1949, Plant Physiol. 24:1-15.

6.2. Results

Preparation of Transgenic Tobacco

A vector was constructed for expression of the VHb gene in plants (Figure 2). Genomic DNA from several tobacco transformants was purified and analyzed by PCR to determine if the VHb gene was inserted into the genome. Approximately 90% of the analyzed plants were positive. PolyA+ mRNA was isolated from plant leaves of two PCR-positive transgenic plant lines, He3 and He4, and a wild-type control line. Two polynucleotide primers, H1 and H2 (see Experimental protocol), were used in a reverse transcriptase PCR reaction to amplify and detect the mRNA transcript of the VHb gene. Both transgenic tobacco lines exhibited a full-length VHb mRNA transcript which corresponded well to the theoretical size of 460 base pairs.

Expression of VHb protein in tobacco could easily be detected by Western immunoblot analysis on leaf and stem protein extracts. A 15.7 kDa protein that cross-reacted with a VHb antiserum was present in both transgenic tobacco lines but absent in the wild-type control. The level of VHb expression using the 35S CaMV promoter of pBi121:2 is approximately 0.1 % of total protein. However, the presence of several cytochromes in tobacco precluded the use of the CO binding assay for VHb activity (Saiki, *et al.*, 1988, Science 239:487-491).

Improved overall growth in transgenic tobacco

Transgenic tobacco seeds of two transgenic lines, He3 and He4, from the F1 generation and a wild-type control were germinated in soil and their development was followed visually and by harvesting plants at regular intervals. Germination was markedly accelerated in plants transformed with VHb-expression constructs, particularly in the transgenic line He4, over that in non-transformed plants. Analysis showed that the germination time for the transformed plants was approximately 3-4 days compared to 6-8 days for wild-type tobacco. Additionally, the growth rates and final yields were higher in the transgenic plants (Figure 3). For instance, the dried weights were enhanced by 80 to 100% after 35 days of

growth. Furthermore, the period between germination and flowering of the transgenic plants was reduced by 3-5 days compared to wild-type.

Alkaloid analysis of the VHb-expressing transgenic tobacco Extracts of two VHb-expressing transgenic tobacco plant lines, He3 and He4, and a wild-type control were analyzed for nicotine and anabasine content by gas chromatography (Figure 4). Compared with the wild-type, the transgenic plaints (He3) were holding on average 34% higher ratio of nicotine corresponding to 108 μg/g dry weight. At the same time, the anabasine levels decreased substantially in the VHb-expressing transgenic plants. The highest yielding individual transgenic tobacco plant had 130 μg/g dry weight of nicotine which is very similar to the highest nicotine content in tobacco given in the literature (Sheen, S.J., 1988, J. Food Sci. 53:1572-1573), 140 μg/g dry weight.

Chlorophyll content of VHb-expressing transgenic tobacco

Visual observation of VHb-expressing tobacco transgenes indicated that many of these plants appeared more green than wild-type control plants. As depicted in Figure 1B, several steps in chlorophyll and heme biosynthesis are oxygen dependent. Therefore, chlorophyll and growth rate assays were conducted on several Fl generation plants obtained from wild-type tobacco and from the VHb-expressing clones designated He3 and He4. Assays of chlorophyll content showed small but significant increases in chlorophyll content particularly in the He4 plants (Figure 5). These plants also grew about 20-40% faster than the wild-type as determined by height. Differences in growth rate between transgenic and wild-type plants were not observed when these plants were grown in an atmosphere with reduced oxygen.

6.3. Discussion

Engineering VHb expression in tobacco plants influenced nicotine (and anabasine) production, but indications of more global effects of VHb expression began with the observation that young transgenic plants exhibit a darker green shade of color than do wild-type plants. Subsequent measurements of chlorophyll levels revealed 10-20% increased levels in the He3 plants which also corresponded 20-40% increased growth rates as determined by plant heights. Cloning of VHb in tobacco affected the He3 and He4 plants somewhat differently. Enhanced levels of

nicotine and chlorophyll were associated with He3 and He4 plants, respectively. Such differences among clones are common, particularly when using integrating vectors.

One hypothesis to explain this dramatic shift in nicotine and anabasine production in VHb-expressing transgenic tobacco plants, is that the presence of VHb increases the oxygen availability, which in turn significantly affects distribution of secondary metabolite production in the whole tobacco plant. Alternatively, these shifts in secondary metabolite production in VHb-expressing transgenes may result from other metabolic effects of VHb such as an increase in the ATP synthesis rate and/or membrane energy. Availability of more energy in the cell might lead to the improvement of many metabolic functions including increased nicotine production.

7. EXAMPLE: EXPRESSION OF VITREOSCILLA HEMOGLOBIN IN DATURA

Hyoscyamine and scopolamine are tropane alkaloids produced by plants; hyoscyamine is the levorotatory component of racemic atropine. Atropine, hyoscyamine, and scopolamine affect the parasympathetic nervous system and exhibit a wide range of pharmaceutical activity. In medicine they are used to relieve Parkinson-ism, dilate pupils, increase the heart rate, counteract toxic agents, and reduce secretions such as sweat. Scopolamine produces fewer undesirable side effects than hyoscyamine and is the preferred substance for treatment of motion sickness and production of derivative drugs for gastric disorders. Therefore, there has been a long-standing interest in raising the scopolamine level of producing plants. *Datura innoxia* a medicinal plant grown in many tropical and sub-tropical countries is one of the principal sources of tropane alkaloids. The biosynthesis of these alkaloids involves several oxygen-requiring steps.

7.1. Experimental Protocol

Recombinant DNA techniques, Transcription analysis and Western immunoblot analysis were performed as above for tobacco. The same transformation vector which was used for tobacco was also used for *Datura*.

Plant material

Plants of *Datura innoxia* Mill were exposed to 16 h light at 24°C and were kept in sterile jars on basal medium (BM) containing Murashige and Skoog's macro salts, and Nitsch and Nitsch micro salts and vitamins (Sangwan *et al.*, 1991, Plant Cell Rep. 10:90-93). The plants were transformed by *Agrobacterium*-mediated gene transfer using the leaf disc method as described by Sangwan *et al.* (1991). Transformants of the F_0 generation were used in this experiment.

Alkaloid analysis

Transgenic *Datura* and wild-type plants were analyzed for their scopolamine and hyoscyamine levels. The plants were grown in soil until they were budding, then the leaves were harvested and dried. The scopolamine and hyoscyamine levels were determined using HPLC as described by Gontier *et al.*, 1994, Agro-Food-Ind. Hi-Tech 5, 26-28.

7.2. Results

Several transgenic plants were positive for the VHb gene as determined by PCR of genomic DNA. PolyA$^+$ mRNA was isolated from plant leaves of six PCR-positive transgenic plants and two wild-type plants. All six transgenes exhibited a full-length VHb transcript. Preliminary results showed that the *Datura* plants transformed with the VHb expression construct also exhibited increased growth rates as compared to the wild-type plants.

Extracts from the six transgenic and two wild-type *Datura* plants were analyzed for scopolamine and hyoscyamine content by HPLC. Results are presented in the following table.

TABLE I

Transgenic plant No.	Production of Scopolamine mg/100 g D.W.	Production of Hyoscyamine mg/100 g D.W.
1	115	47
2	148	35
3	246	20
4	86	28
5	118	17
6	70	10
Control A	26	7.1
Control B	55	8.0

7.3. Discussion

Similarly to the results in tobacco, engineering *Datura* plants to express a VHb gene also resulted in transformed plants which both grew faster and contained altered levels of desired secondary metabolites. These results show that the expression of oxygen-binding proteins in whole plants, particularly proteins which reversibly bind oxygen via a heme group such as VHb, is generally applicable to obtain improved transformed plants.

8. EXAMPLE: PHENOTYPIC SCREEN FOR VITREOSCILLA HEMOGLOBIN MUTANTS

In order to create a library of randomly mutated VHbs, we used an earlier construct, pINT1 which carries the VHb gene as a template (Khosla and Bailey, 1989, *supra.*). The VHb sequence used as a template encoded the wild-type VHb protein with an additional 8 amino acid extension - MTMITPSF - at the amino-terminal end of the protein. However, all numbering conventions given herein correspond to the wild-type amino acid sequence positions as described in U.S. Patent No. 5,049,493 *supra.*. The VHb gene was amplified by low fidelity PCR with an estimated error rate of 2-4% (Leung *et al.*, 1989, Technique, 1, 11-15) using primers H1 and H2.

8.1. Experimental Protocol

PCR

Low fidelity PCR was performed with 1 mM dCTP, dGTP and dTTP, 0.2 mM dATP, 0.45 mM of each primer, 0.1 ng template, 1% 1 M mercaptoethanol, 10% DMSO, 0.5 mM $MnCl_2$ and 10 U Taq DNA polymerase. A temperature profile was employed of 2 min. at 95°C, hold at 70°C while the Taq DNA polymerase was added, 30 cycles of 15 sec. at 92°C, 30 sec. at 50°C, 1 min. at 72°C and an additional 5 cycles of 15 sec. at 92°C, 5 min. at 72°C. The amplified PCR fragments were inserted into pUC19 which was subsequently transformed into *E coli*. Sequencing of the mutated genes was performed on an automatic sequencer with the terminator chemistry.

Screening the library for colonies carrying mutated VHb

The plasmid library was transformed into *E. coli*. TG1 and selected on LB agar (Sambrook *et al.*) by visually assessing the color of the colonies. Expression of wild-type VHb results in a red colored colony. The red color phenotype is indicative of a VHb protein capable of binding oxygenated heme. Approximately 10,000 colonies were identified ranging in color from white to deep red.

Western immunoblot analysis

Overnight cultures were pelleted, resuspended in a sodium phosphate buffer (50 mM, pH 7.0) and sonicated to obtain protein extracts. The proteins were separated on 15% SDS-PAGE gels and blotted onto Immobilon P nylon membranes supplied by Millipore. The electrophoresis and blotting were performed as described by Laemmli (1970) and Sambrook *et al.* (1989), respectively. The VHb mutants were detected by VHb antiserum generated in rabbit, and visualized by peroxidase conjugated swine anti-rabbit immunoglobulins (Dako).

8.2. Results

Randomly picked colonies, which exhibited the correct size VHb (15.7 kDa) according to Western blot, were sequenced. The following altered VHb amino acid sequences were identified.

TABLE II.

ALTERED VHB PROTEINS OBTAINED THROUGH PCR MUTAGENESIS

ISOLATE	AMINO ACID ALTERATION
A	$Ile^{129}Thr$
B	$His^{36}Arg$
C	$Lys^{79}Asn$
D	$Phe^{33}Tyr$, $Gln^{53}Arg$
E	$Lys^{124}Glu$, $Phe^{133}Leu$, $Ile^{134}Thr$
F	$Gly^{21}Asp$, $Val^{136}Glu$, $Ala^{138}Thr$
G	$Leu^{51}Ser$, $Val^{83}Ala$, $Tyr^{126}His$
H	$Lys^{11}Glu$, $Lys^{107}Glu$, $Lys^{124}Stop$

8.3. Discussion

The PCR mutagenesis technique allowed the phenotypic selection of VHb mutants which retain at least some functional activity of the VHb, as evidenced by their ability to bind oxygenated heme. Although these mutants were generated in a VHb protein with an 8 amino acid amino-terminal extension, these mutants will be similarly effective in a VHb protein with the wild-type amino-terminus. Further, this approach will enable the easy selection and generation of any altered hemoglobin molecule through mutagenesis of the coding sequence, followed by selection for clones which encode heme containing proteins which retain functional activity.

9. EXAMPLE: TRANSFORMATION OF RICE WITH A VITREOSCILLA HEMOGLOBIN EXPRESSION CONSTRUCT

This example illustrates the transformation of rice using particle bombardment with a polynucleotide that directs the expression of Vitreoscilla hemoglobin in plant cells.

9.1 Experimental Protocol

Plasmid constructions

pJFK2 was constructed as a derivative of pUC that contains the *vhb* gene under the regulation of the CaMV 35S promoter and a nos polyadenylation

sequence. Additionally, the stop codon of the *vhb* gene was altered using PCR mediated mutagenesis from TAA to TGA (a more frequently used stop codon in plants). For the selection of transgenic rice tissue, the *aph*IV gene from *E.coli* which confers resistance to Hygromycin was included in the transformation construct. This marker gene was under the control of the CaMV 35S promoter and the CaMV polyadenylation sequence. A map of pJFK is shown in Figure 6.

The pJFK2 plasmid was digested with I-SceI, and the larger fragment containing *vhb* gene and hygromycin resistance gene (and lacking the ampicillin resistance gene) was isolated for use in posterior bombardment of plant embryos.

Rice transformation

Immature *Oryza sativa* embryos of the Japonica variety Taipei 309 were aseptically isolated 10-14 days after pollination from green-house-plants and plated scutulum site up on solid MS medium (Murashige and Skoog, 1962) containing 3% sucrose, 2 mg l^{-1} 2,4-D and 50 mg l^{-1} cefotaxime (MS1). After 4-6 days (28°C, darkness) embryos were transferred to solid MS medium containing 10% sucrose, 2 mg l^{-1} 2,4-D and 50 mg l^{-1} cefotaxime (MS2) and subjected within 1 hour to microprojectile bombardment with a particle inflow gun. The DNA fragment containing *vhb* and *aph* IV genes (5μg) was precipitated on 1-3 μm gold particles (Aldrich) as described by Vain et al., 1993, Plant Cell Rep. 12:84-88. Gold particles (400 μg per bombardment) were accelerated to the target with a particle inflow gun (Finer et al., 1992, Plant Cell Rep. 11:323-328) at a pressure of 6 bar. Embryos were placed 16 cm below the syringe filter. Twenty-four hours post-bombardment embryos were subjected to selection on a solid selection media (MS1 medium containing 20 mg l^{-1} hygromycin B) and incubated at 28°C in the dark.

After 1 week embryos were transferred to a liquid selection media, R2 medium (Ohira et al., 1973, Plant Cell Physiol. 14:113-121) supplemented with: 3% sucrose, 1 mg l^{-1} thiamine, 1 mg l^{-1} 2,4-D, 50 mg l^{-1} cefotaxime and 20 mg l^{-1} hygromycin B. The embryos were incubated with shaking at 28°C in the dark and subcultured weekly. Developing calli were isolated 3 to 6 weeks later, and transferred to a callus increasing media (R2 medium supplemented with: 6% sucrose, MS vitamins, 100 mg l^{-1} inositol, 2 mg l^{-1} 2,4-D, 50 mg l^{-1} cefotaxime and 20 mg l^{-1} hygromycin B). The calli were incubated in this media at 28°C in the dark and subcultured weekly.

Resistant calli were transferred to solid R2 regeneration medium supplemented with 2% sucrose, 3% sorbitol, 20 mg l^{-1} Hygromycin B, 1 mg l^{-1} of zeatin, 0.5 mg l^{-1} IAA, MS vitamins and 0.65% agarose. The callus tissue was maintained at 28°C with 12 h of light in order to enhance shoot formation. The calli were then subcultured every 3 weeks until shoots had reached a length of 2-3 cm. They were transferred to half-strength MS rooting medium without hormones, supplemented with 1.5% sucrose and 0.3% gelride®. After 2-4 weeks of cultivation, plantlets were transferred directly to the greenhouse and planted in soil. Plantlets were grown in 7 liter aquaculture pots with fertilizer enriched earth, 3 plants per pot, (day: 12 h, 28°C, 80% humidity; night: 12h, 21°C, 60% humidity) until they flowered and set seeds.

Analysis of transgenic rice plants

To check for presence of the transgene, complexity of insertion(s) and number of copies present, Southern blot analysis was performed as described previously (Burkhardt et al., 1997). A PCR-amplified, DIG-labeled (Boehringer), 300-bp fragment of the coding region of *vhb* was used as a probe.

Transgene expression is confirmed using western immunoblots as described above and in the literature (Wünn et al., 1996, Bio/Technology 14:171-176).

9.2 Results

Transformed plants were analyzed by Southern blot and were found to carry the transgene. Western blot is performed and demonstrates that the transformed plants express the *vhb* transgene. Seeds from the transgenic R0 plants are collected and R1 plants are grown. Growth of the R1 generation plants, as compared to control plants without the transgene, is more rapid.

Although the invention is described in detail with reference to specific embodiments thereof, it will be understood that variations which are functionally equivalent are within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings such modifications are intended to fall within the scope of the appended claims. Various publications are cited herein, each of the disclosures of which is incorporated by reference in its entirety.

WHAT IS CLAIMED IS:

1. A plant which has been genetically engineered by transformation with a polynucleotide encoding a globin protein.

2. The plant of claim 1, wherein the globin protein is selected from the group consisting of *Vitreoscilla* hemoglobin, horse heart myoglobin, E. coli hemoprotein, B. subtilis hemoprotein, yeast flavohemoglobin, soybean leghemoglobin, lupin leghemoglobin, and sperm whale myoglobin, or their functional equivalents.

3. The plant of claim 1 or 2, wherein the polynucleotide encoding the globin protein is operably linked to a plant promoter.

4. The plant of claim 3, wherein the plant promoter is a strong, constitutively expressed plant promoter.

5. The plant of claim 4, wherein said strong, constitutively expressed plant promoter is a CaMV 35S promoter.

6. The plant of claim 3, wherein the plant promoter is a tissue specific plant promoter.

7. The plant of any one of claims 1-6, wherein the polynucleotide is introduced into a plant cell via microprojectile bombardment.

8. The plant of any one of claims 1-6, wherein the polynucleotide is introduced into a plant cell on a Ti plasmid.

9. The plant of any one of claims 1-8, wherein the plant is a dicotyledon.

10. The plant of claim 9, wherein the plant is from the genus *Nicotiana*.

11. The plant of claim 10, wherein the plant is *Nicotiana tabaccum*.

12. The plant of claim 9, wherein the plant is from the genus *Datura*.

13. The plant of any one of claims 1-8, wherein the plant is a monocotyledon.

14. The plant of claim 13, wherein the plant is from the genus *Oryza*.

15. A method of producing a plant with an improved agronomic or medicinal characteristic comprising transforming the plant with a polynucleotide encoding a globin protein, wherein the transformed plant exhibits an increased rate of growth, an increase in vegetative yield of the mature plant, an increased rate of germination, and/or an increased chlorophyll content, relative to a non-transformed plant.

16. The method of claim 15, wherein the globin protein is selected from the group consisting of *Vitreoscilla* hemoglobin, horse heart myoglobin, E. coli hemoprotein, B. subtilis hemoprotein, yeast flavohemoglobin, soybean leghemoglobin, lupin leghemoglobin, and sperm whale myoglobin, or their functional equivalents.

17. The method of claim 15 or 16, wherein the polynucleotide encoding the globin protein is operably linked to a plant promoter.

18. The method of claim 17, wherein the plant promoter is a strong, constitutively expressed plant promoter.

19. The method of claim 18, wherein the strong, constitutively expressed plant promoter is a CaMV 35S promoter.

20. The method of claim 17, wherein the plant promoter is a tissue specific plant promoter.

21. The method of any one of claims 15-20, wherein the polynucleotide is introduced into a plant cell via microprojectile bombardment.

22. The method of any one of claims 15-20, wherein the polynucleotide is transformed into a plant cell on a Ti plasmid.

23. The method of any one of claims 15-22, wherein the plant is a dicotyledon.

24. The method of claim 23, wherein the plant is from the genus *Nicotiana*.

25. The method of claim 24, wherein the plant is *Nicotiana tabaccum*.

26. The method of claim 25, wherein the plant further exhibits an elevated level of nicotine.

27. The method of claim 23, wherein the plant is from the genus *Datura*.

28. The method of claim 27, wherein the plant further exhibits an elevated level of scopolamine and/or hyoscyamine.

29. The method of claim 17-22, wherein the plant is a monocotyledon.

30. The plant of claim 29, wherein the plant is from the genus *Oryza*.

31. A plant produced by the method of claim 17.

32. A method of altering the distribution of secondary metabolites in a plant, comprising transforming a target plant or target plant cell with a polynucleotide encoding a globin protein operably linked to a plant promoter.

33. The method of claim 32, wherein the plant is from the genus *Nicotiana* and the distribution of nicotine is altered relative to anabasine.

34. The method of claim 32, wherein the plant is from the genus *Datura* and the distribution of scopolamine and/or hyoscyamine is altered.

35. A mutant *Vitreoscilla* hemoglobin protein, wherein the mutant *Vitreoscilla* hemoglobin protein differs from the wild type *Vitreoscilla* hemoglobin protein by having an amino acid substitution selected from the group consisting of: $Ile^{129}Thr$; $His^{36}Arg$; $Lys^{79}Asn$; $Phe^{33}Tyr$, $Gln^{53}Arg$; $Lys^{124}Glu$, $Phe^{133}Leu$, $Ile^{134}Thr$; $Gly^{21}Asp$, $Val^{136}Glu$, $Ala^{138}Thr$; $Leu^{51}Ser$, $Val^{83}Ala$, $Tyr^{126}His$; and $Lys^{11}Glu$, $Lys^{107}Glu$, $Lys^{124}Stop$.

36. The mutant *Vitreoscilla* hemoglobin protein of Claim 35 which further comprises an amino-terminal amino acid extension having the sequence MetThrMetIleThrProSerPhe.

37. A polynucleotide that encodes the mutant *Vitreoscilla* hemoglobin protein of Claim 35 or 36.

38. A host cell that contains the polynucleotide of Claim 37.

39. The host cell of Claim 38, wherein the host cell is selected from the group consisting of a plant cell, a bacterial cell, an animal cell, a fungal cell and an algae cell.

40. A method of generating a polynucleotide encoding mutant *Vitreoscilla* hemoglobin protein comprising:

amplifying a polynucleotide that encodes *Vitreoscilla* hemoglobin protein using a low fidelity polymerase chain reaction;

transforming the amplified polynucleotides into bacteria; and

selecting bacteria that contain a polynucleotide sequence encoding a mutant *Vitreoscilla* hemoglobin protein.

41. The method of Claim 40, wherein said selecting step comprises assaying the color of the bacterial cells.

42. A polynucleotide that encodes a mutant *Vitreoscilla* hemoglobin protein isolated by the method of Claim 40 or Claim 41.

43. A mutant *Vitreoscilla* hemoglobin protein encoded by the polynucleotide of Claim 42.

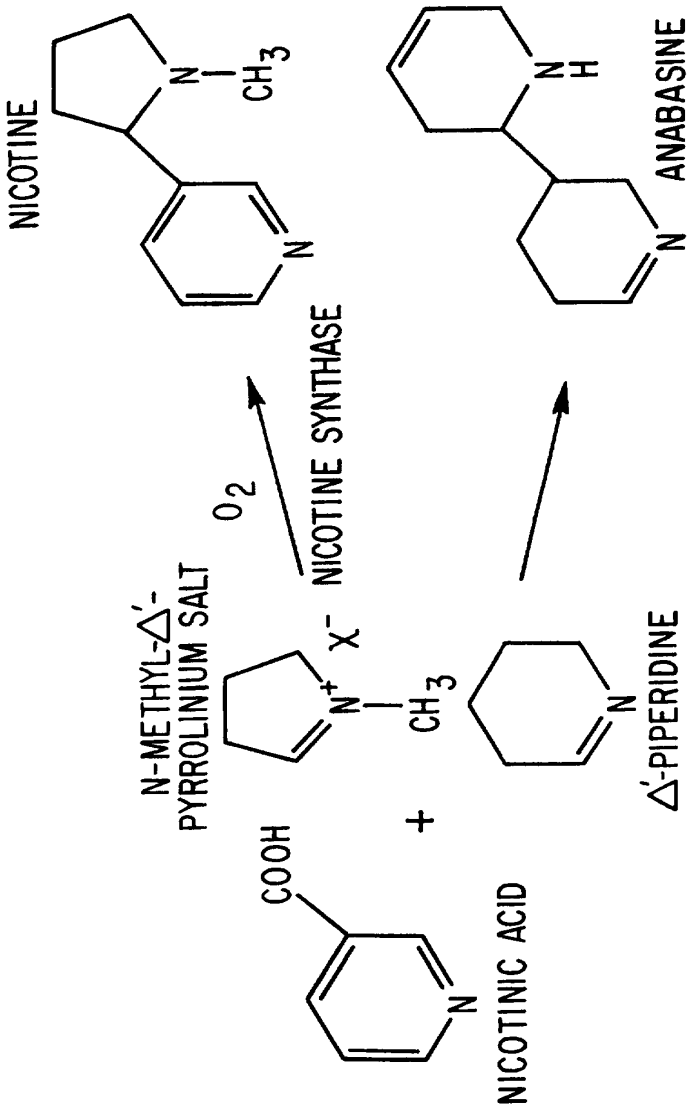

NICOTINE
ANABASINE
N-METHYL-Δ'-PYRROLINIUM SALT
O_2
NICOTINE SYNTHASE
X^-
CH_3
Δ'-PIPERIDINE
COOH
NICOTINIC ACID

FIG. 1A

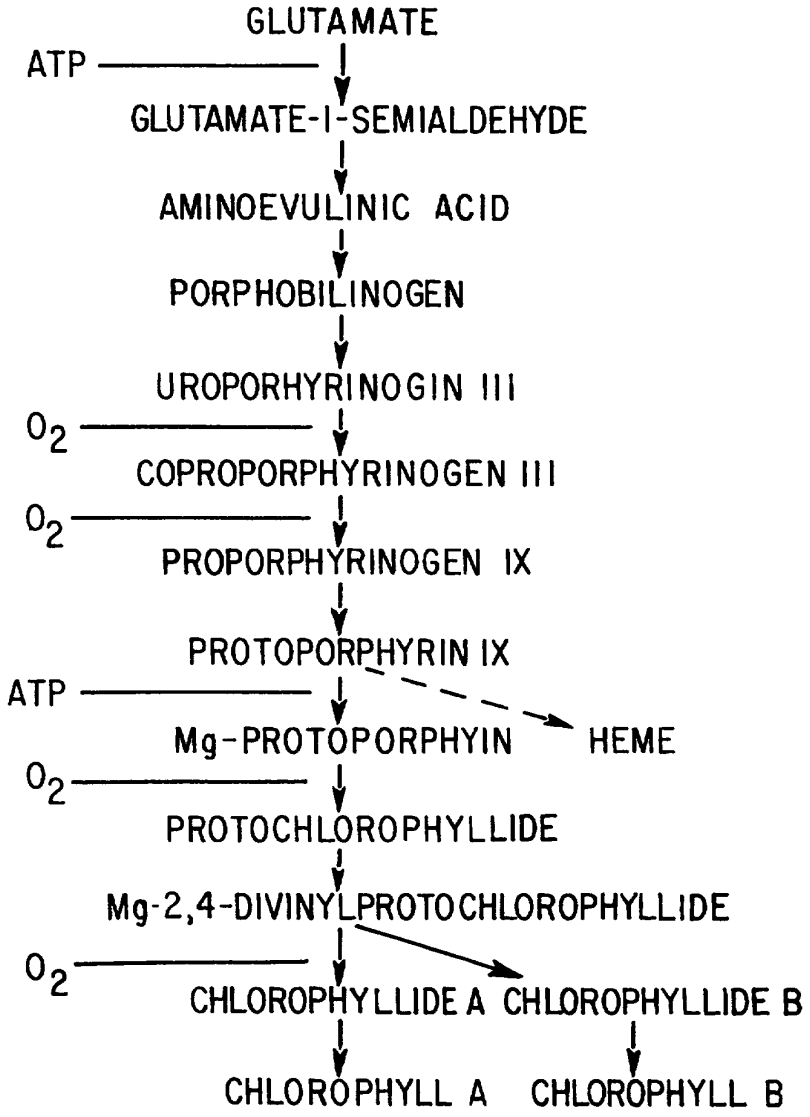

GLUTAMATE
ATP
GLUTAMATE-I-SEMIALDEHYDE
AMINOEVULINIC ACID
PORPHOBILINOGEN
UROPORHYRINOGIN III
O_2
COPROPORPHYRINOGEN III
O_2
PROPORPHYRINOGEN IX
PROTOPORPHYRIN IX
ATP
Mg-PROTOPORPHYIN
HEME
O_2
PROTOCHLOROPHYLLIDE
Mg-2,4-DIVINYLPROTOCHLOROPHYLLIDE
O_2
CHLOROPHYLLIDE A
CHLOROPHYLLIDE B
CHLOROPHYLL A
CHLOROPHYLL B

FIG. 1B

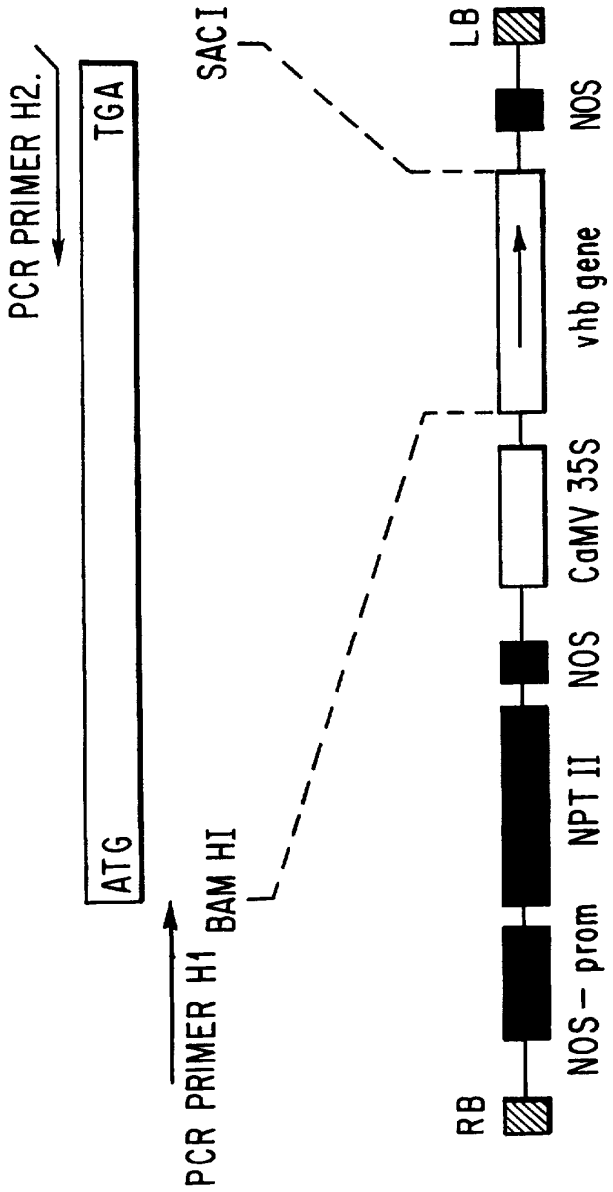

PCR PRIMER H2.
TGA
ATG
PCR PRIMER H1
BAM HI
SACI
LB
NOS
vhb gene
CaMV 35S
NOS
NPT II
NOS – prom
RB

FIG. 2

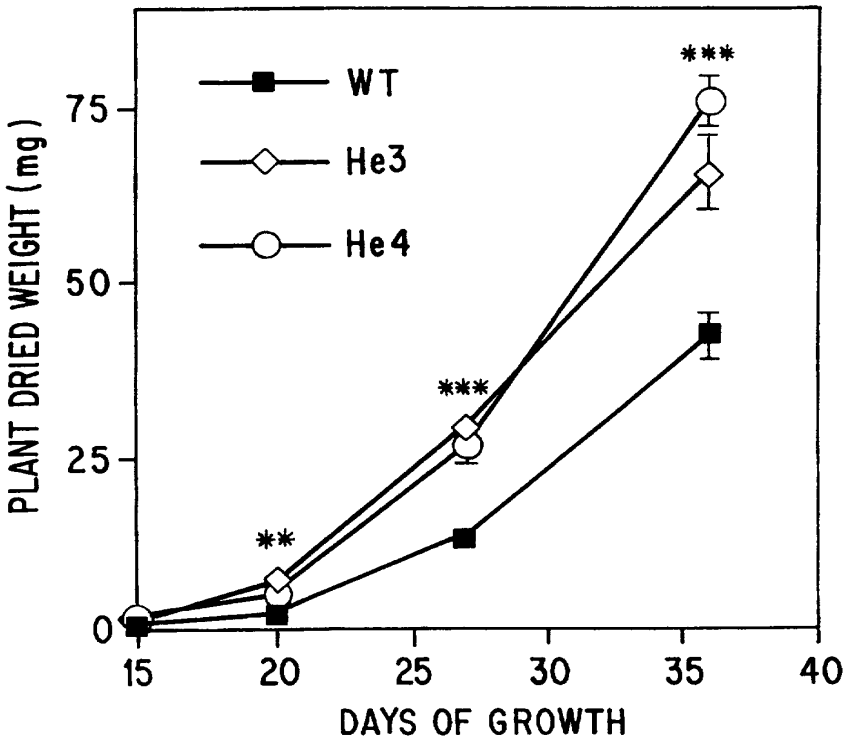

WT
He3
He4
***
***
**
PLANT DRIED WEIGHT (mg)
75
50
25
0
15
20
25
30
35
40
DAYS OF GROWTH

FIG. 3

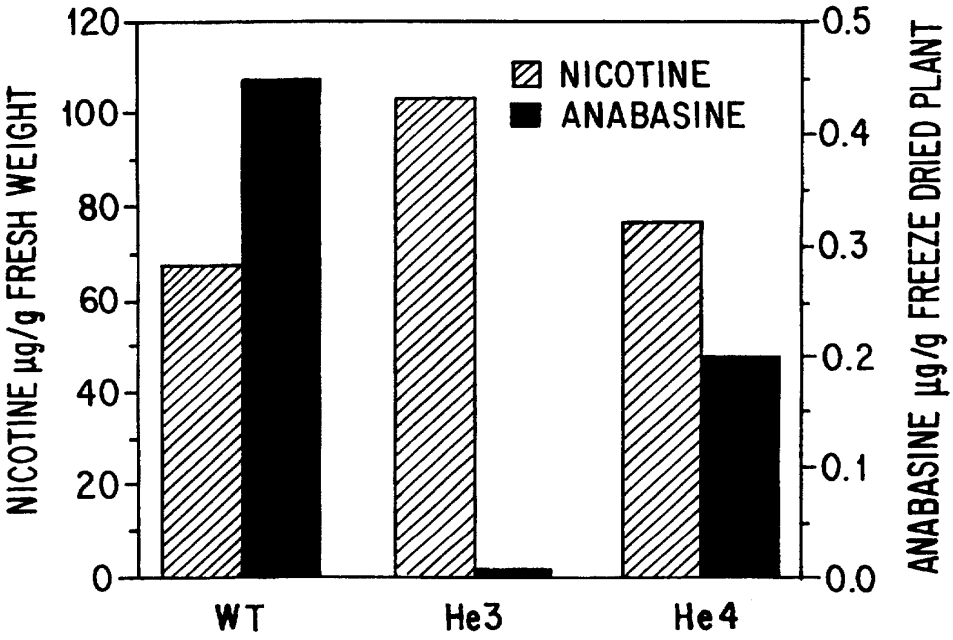

NICOTINE
ANABASINE
NICOTINE µg/g FRESH WEIGHT
ANABASINE µg/g FREEZE DRIED PLANT
120
100
80
60
40
20
0
0.5
0.4
0.3
0.2
0.1
0.0
WT
He3
He4

FIG. 4

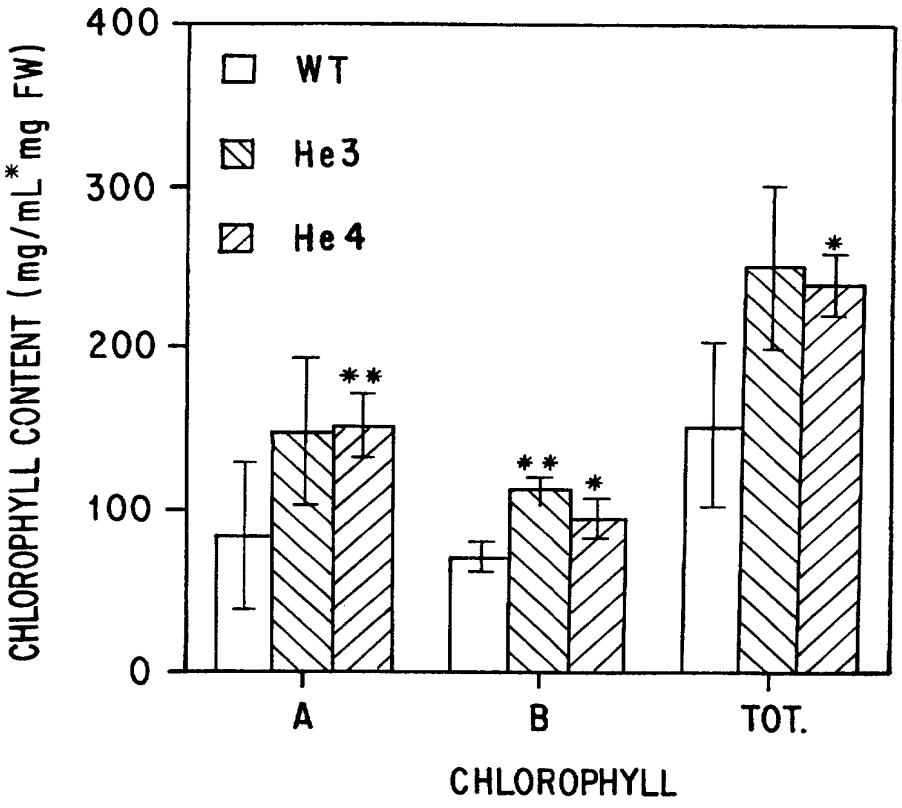

400
300
200
100
0
CHLOROPHYLL CONTENT (mg/mL*mg FW)
WT
He3
He4
**
**
*
*
A
B
TOT.
CHLOROPHYLL

FIG. 5

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

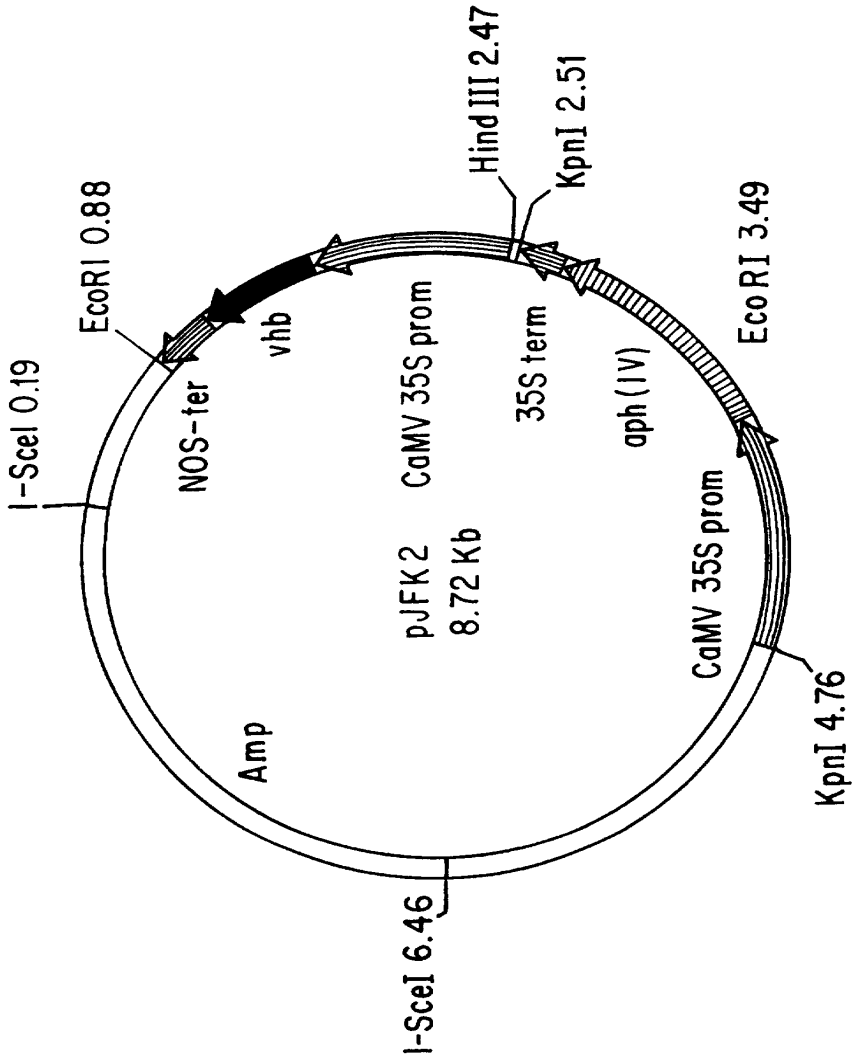


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17246

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :A01H 5/00; C12N 5/14, 15/31, 15/82
US CL :435/172.3, 320.1, 419; 536/23.7; 800/205, DIG40, DIG43, DIG57
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/172.3, 320.1, 419; 536/23.7; 800/205, DIG40, DIG43, DIG57

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS, STN, BIOSIS, EMBASE, CA, WPIDS
oygen binding, hemoglobin, plant, transgenic, transformation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DE ZOETEN et al. The Expression, Localization, and Effect of a Human Interferon in Plants. Virology. 1989, Vol. 172, pages 213-222, see entire article.	1-6, 1-12, 14-20, 24-28, 31-34
Y	KHOSLA et al. Heterologous Expression of a Bacterial Haemoglobin Improves the Growth Properties of Recombinant Escherichia coli. Nature. 18 February 1988, Vol. 331, pages 633-635, see entire article.	1-6, 10-12, 14-20, 24-28, 31-34
Y	BOGUSZ et al. Nonlegume Hemoglobin Genes Retain Organ-Specific Expression in Heterologous Transgenic Plants. The Plant Cell. July 1990, Vol. 2, pages 633-641, see abstract.	1-6, 10-12, 14-20, 24-28, 31-34

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 28 OCTOBER 1997	Date of mailing of the international search report 12 DEC 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer AMY NELSON Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17246

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 7, 8, 9, 13, 21-23, 29-30
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-6, 10-12, 14-20, 24-28, 31-34

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17246

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1.

Group I, claim(s)1-6, 10-12, 14-20, 24-28, 31-34, drawn to method of producing transgenic plants comprising a globin gene, and transgenic plants thereby obtained.
Group II, claim(s) 35-36 and 43, drawn to mutant hemoglobin protein.
Group III, claim(s) 37-42, drawn to method of making mutant hemoglobin gene and mutant hemoglobin gene thereby obtained.

The inventions listed as Groups I, II and III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Group I is drawn toward wild type globin genes and Group III is drawn toward mutant globin genes and hence they lack a special technical feature.

Both Group III encompasses many different mutant genes and hence the mutant genes of Group III are not linked by a special technical feature to the proteins of Group II.

DNA and proteins have completely different structures, compositions and functions, and mutant DNA molecules and proteins differ structurally and functionally from wild type DNA and proteins. Hence, the inventions of Group I, II and III are not so linked as to form a single inventive concept under PCT Rule 13.1.

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
15. Juli 2004 (15.07.2004)

PCT

(10) Internationale Veröffentlichungsnummer
WO 2004/057946 A2

(51) Internationale Patentklassifikation[7]: A01H 1/00

(21) Internationales Aktenzeichen: PCT/EP2003/014774

(22) Internationales Anmeldedatum:
23. Dezember 2003 (23.12.2003)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
102 60 707.9 23. Dezember 2002 (23.12.2002) DE

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: **MAX-PLANCK-GESELLSCHAFT** [DE/DE]; Hofgartenstr. 8, 80539 München (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: **GEIGENBERGER, Peter** [DE/DE]; Quantzstr. 12, 14129 Berlin (DE). **LANGER, Anke** [DE/DE]; Sellostr. 27, 14471 Potsdam (DE). **VIGEOLAS, Helene** [FR/DE]; Am alten Mörtelwerk 14, 14469 Potsdam (DE). **STITT NIGEL, Marc** [DE/DE]; Grosse Weinmeisterstr. 22a, 14469 Potsdam (DE). **THOMAS VAN DONGEN, Joost** [NL/DE]; Leipziger Str. 52, 14473 Potsdam (DE). **UDVARDI, Michael** [DE/DE]; Golmer Fichten 32, 14476 Golm (DE).

(74) Anwalt: **FITZNER, Uwe**; Lintorfer Strasse 10, 40878 Ratingen (DE).

(81) Bestimmungsstaaten *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Bestimmungsstaaten *(regional)*: ARIPO Patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:
— *ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts*

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

WO 2004/057946 A2

(54) **Title:** METHOD FOR ALTERING THE CONTENT OF RESERVE SUBSTANCES IN PLANTS

(54) **Bezeichnung:** VERFAHREN ZUR VERÄNDERUNG DES GEHALTS VON SPEICHERSTOFFEN IN PFLANZEN

(57) **Abstract:** The invention relates to a method for altering the content of reserve substances in plants during which leghemoglobin and/or hemoglobin expressing transformed plants are used. The invention also relates to corresponding plants and to the use thereof.

(57) **Zusammenfassung:** Die Erfindung betrifft ein Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen, wobei Leghemoglobin und/oder Hemoglobin exprimierende transformierte Pflanzen eingesetzt werden, entsprechende Pflanzen sowie deren Verwendung.

Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen

Die Erfindung betrifft ein Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen durch Einsatz von Leghemoglobin und/oder Hemoglobin exprimierender transformierter Pflanzen, entsprechende Pflanzen sowie deren Verwendung.

Speicherstoffe in Pflanzen dienen als Reservestoffe und werden von pflanzlichen Geweben gebildet und intrazellulär abgelagert. Als Speicherstoffe dienen z. B. Polysaccharide (Kohlenhydrate wie Lichenin, Stärke, Glykogen, Polyfructosane), Proteine, Fette, auch Polyphosphate und Polyhydroxyalkanoate. Die Speicherstoffe können bei Bedarf wieder in den Stoff- und Energiewechsel eingeschleust werden, beispielsweise bei Nahrungsmangel, der Keimung von Samen, dem Wachstum und sonstigen energieverbrauchenden Vorgängen.

Speicherstoffe sind wertvolle Rohstoffe für die Ernährung (Getreide, Nüsse, Früchte, Obst, Gewürze etc.) und bilden die Basis vieler menschlicher Nahrungsmittel und können auch als Lieferanten für technische Fette und Öle dienen. Andere Speicherstoffe werden wegen ihrer pharmakologisch wirksamen Inhaltsstoffe arzneilich genutzt. (Quelle: CD Römpp Chemie Lexikon – Version 1.0, Stuttgart/New York: Georg Thieme Verlag 1995).

Auch werden in zunehmendem Maße die Speicherstoffe als nachwachsende und ökologisch verträgliche Rohstoffe für die Industrie angewendet, wie dies z. B. durch die Verwendung von Stärke zur Herstellung von Verpackungsmaterialien oder z. B. von Pflanzenölen als Kraftstoffe, wie Biodiesel oder Schmierstoffe möglich ist.

Die sogenannten sekundären Stoffwechselprodukte (sek. Metaboliten, vgl. a. Stoffwechsel) von Pflanzen (z. B. Farbstoffe, Gifte, etherische Öle, Alkaloide, Fruchtsäuren) werden im allg. nicht zu den Speicherstoffen gezählt (Quelle: Römpp Lexikon Chemie – Version 2.0, Stuttgart/New York: Georg Thieme Verlag 1999 "Reservestoffe").

In Pflanzen werden die Speicherstoffe in Samen oder Speicherorganen von Kohlenhydrat-Vorstufen gebildet. Dabei ist Saccharose die Primärquelle von Kohlenstoff und Energie, welche von den Blättern in die sich entwickelnden Samen bzw. in Speicherorgane transportiert wird. Die Saccharose in den Blättern wird dabei durch die aus der Photosynthese gewonnene Stärke gebildet. Die Saccharose, die in die sich entwickelnden Samen transportiert wird, dient neben der Synthese von Fettsäuren für die Speicherlipide auch zur Synthese von Speicherproteinen und Speicherstärke.

Insgesamt enthalten z. B. Samen drei verschiedene Formen an Speicherstoffen: Speicherlipide, Speicherproteine und Stärke. Je nach Pflanzen variieren dabei die Verhältnisse der drei Speicherstoffe zueinander. So enthalten Rapssorten z. B. etwa 48 % Speicherlipide, 19 % Stärke und 21 % Speicherproteine, während Sojabohne 22 % Lipide, 12 % Stärke und 37 % Proteine enthält (Biochemistry & Molecular Biology of the Plant ed. Buchanan, Gruissem, Jones 2000, American Society of Plant Physiologists) bezogen auf ihre Trockenmasse.

Die aus den pflanzlichen Ölen (Lipide) erhältlichen Fettsäuren sind von besonderem Interesse. Sie kommen beispielsweise als Grundstoffe für Weichmacher, Schmierstoffe, Tenside, Kosmetika usw. zum Einsatz oder werden in der Lebens- und Futtermittelindustrie als wertvolle Grundstoffe eingesetzt. So ist beispielsweise die Bereitstellung von Rapsölen mit

Fettsäuren mittlerer Kettenlänge von besonderem Interesse, da diese besonders für die Tensidherstellung begehrt sind.

Durch die gezielte Modulation pflanzlicher Stoffwechselwege mittels gentechnischer Verfahren kann der pflanzliche Metabolismus in einer Weise vorteilhaft verändert werden, die durch klassische Züchtungsmethoden nur über langwierige Schritte bzw. überhaupt nicht zu erreichen wären. So werden z. B. ungewöhnliche Fettsäuren, beispielsweise bestimmte mehrfach ungesättigte Fettsäuren, nur in bestimmten Pflanzen bzw. überhaupt nicht in Pflanzen synthetisiert und können deshalb nur durch Expression des entsprechenden Enzyms in transgenen Pflanzen gezielt hergestellt werden (z. B. Millar et al. (2000) Trends Plant Sci 5:95-101).

Stärke als Speicherstoff kann nicht nur in Samen eingelagert werden, sondern auch in anderen Speicherorganen. Wichtige Speicherorgane für Stärke sind dabei Hypokotyl und Wurzeln. Durch Vermehrung und Vergrößerung der Zellen des Rindenparenchyms entstehen Wurzelknollen, wie z. B. Kartoffelknollen oder bei Verdickung des Wurzelhalses Wurzelrüben, wie z. B. Zuckerrübe oder Yams.

So ist Stärke z. B. der Hauptbestandteil der Kartoffel-Trockensubstanz. Neben der Verwendung als Lebensmittel dient die Kartoffel daher auch als Futtermittel und als Rohstoff zur Gewinnung von Stärke und Alkohol. Weltweit wurden 1988 269,7 Mio. t Kartoffeln geerntet. (Quelle: CD Römpp Chemie Lexikon – Version 1.0, Stuttgart/New York: Georg Thieme Verlag 1995).

Die Patente US 6,372,961, WO98/12913 und WO00/00597 beschreiben die Nutzung von Hemoglobin oder strukturverwandter Proteine (Myoglobin, bakterielles Hemoglobin) zur Erhöhung der Sauerstoffassimilation in

Pflanzen. Damit soll die Keimung bzw. der Energiehaushalt der entsprechend transgenen Pflanzen verbessert sein auch unter Sauerstoffmangelbedingungen ein normales Wachstum gewährleistet werden. Auch sollen die Menge produzierter sekundärer Metabolite erhöht werden (WO98/12913 Seite 6, Zeile 24). Die Steigerung der Produktion von Speicherstoffen geht aus diesen Schriften nicht hervor.

Aufgabe der vorliegenden Erfindung war es daher, den Gehalt von Speicherstoffen in Speicherorganen von Pflanzen zu erhöhen, um somit eine bessere Nutzung von Anbauflächen, Dünger usw. und eine bessere Ausbeute (oder Ertrag) mittels der Pflanzen zu erwirtschaften. Insbesondere soll die Produktion von Stärke oder Öl verbessert bzw. ermöglicht werden.

Diese Aufgabe wird dadurch gelöst, daß im erfindungsgemäßen Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen Leghemoglobin exprimierende transformierte Pflanzen eingesetzt werden. Die Aufgabe wird ebenfalls durch Leghemoglobin exprimierende transformierte Pflanzen sowie deren Verwendung gelöst.

Überraschenderweise konnte festgestellt werden, dass durch Expression eines Leghemoglobins Speicherstoff-haltige transformierte Pflanzen produziert werden, die auf Grund ihres (höheren) Gehalts von Speicherstoffen eine bessere Nutzung von Anbauflächen, Dünger usw. erlauben und daher einen besseren Ertrag der Speicherstoffe, insbesondere Stärke und Öl erlauben. Somit ist eine wirtschaftlich interessante Verwendung der erfindungsgemäßen Pflanzen möglich.

Das Leghemoglobin gehört zur Familie der Hemoglobin-Proteine, deren Funktion die reversible Sauerstoffbindung und Versorgung ist. Es stammt aus

Wurzelknöllchen von Hülsenfrüchten (Leguminosen) und ist eine isolierbare rote Substanz, die dem Myoglobin der Wirbeltiere ähnelt. Durch die reversible Bindung von O_2 kann Leghemoglobin den hohen Sauerstoff-Bedarf bei der Stickstoff-Fixierung durch die Knöllchenbakterien gewährleisten. Das Apoprotein wird von den Pflanzenzellen und das Häm von den Bakterien gebildet (Quelle: CD Römpp Chemie Lexikon – Version 1.0, Stuttgart/New York: Georg Thieme Verlag 1995).

In einer weiteren bevorzugten Variante der Erfindung werden in dem erfindungsgemäßen Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen Hemoglobin oder Leghemoglobin und Hemoglobin exprimierende transformierte Pflanzen eingesetzt. Demgemäß sind Gegenstand der Erfindung auch noch Leghemoglobin und/oder Hemoglobin exprimierende transformierte Pflanzen sowie deren Verwendung.

Unter Hemoglobin werden erfindungsgemäß Eisen-II-Komplexe des Protoporphyrins verstanden.

Unter Expression wird in der vorliegenden Anmeldung die Übertragung einer genetischen Information ausgehend von DNA oder RNA in ein Genprodukt (Polypeptid oder Protein, hier Leghemoglobin oder Hemoglobin) verstanden und soll auch den Begriff der Überexpression beinhalten, womit eine verstärkte Expression gemeint ist, so dass das Fremdprotein oder das natürlich vorkommende Protein verstärkt produziert werden oder einen großen Teil des gesamten Protein-Gehaltes der Wirtszelle ausmachen.

Unter Transformation wird die Übertragung einer genetischen Information in einen Organismus, insbesondere Pflanze verstanden. Darunter sollen alle dem Fachmann bekannten Möglichkeiten zur Einschleusung der Information

fallen, z. B. Mikroinjektion, Elektroporation, Teilchenbeschuss (Partikelbombardement), Agrobakterien oder Chemikalien vermittelte Aufnahme (z. B. Polyethylenglycol-vermittelter DNA-Aufnahme oder über die Siliziumcarbonatfaser-Technik). Die genetische Information kann z. B. als DNA, RNA, Plasmid und auf sonstige Weise in die Zellen eingebracht und entweder durch Rekombination ins Wirtsgenom eingebaut, in freier Form oder unabhängig als Plasmid vorliegen. Eine transformierte Pflanze im Sinne der Erfindung ist also eine gentechnisch veränderte Pflanze.

Unter Speicherstoffe werden Polysaccharide, vorzugsweise Kohlenhydrate, Proteine, Fette, Polyphosphate und Polyhydroxyalkanoate, besonders bevorzugt Lichenin, Stärke, Glykogen, Polyfructosane verstanden.

Unter den aufgezählten Verbindungen werden Kohlenhydrate und Fette ganz besonders bevorzugt. Höchst bevorzugt als Speicherstoffe sind Stärke und Öl.

Stärke ist dem Fachmann bekannt und es wird für weitere Informationen auf das Römpp Chemie Lexikon – CD Version 2.0, Stuttgart/New York: Georg Thieme Verlag 1999, verwiesen.

"Öl" umfasst im Sinne der Erfindung neutrale und/oder polare Lipide und Mischungen derselben. Beispielhaft jedoch nicht einschränkend seien die in Tabelle 1 aufgeführten zu nennen.

Tabelle 1: Pflanzliche Lipidklassen

Neutrale Lipide	Triacylglycerol (TAG)
	Diacylglycerol (DAG)
	Monoacylglycerol (MAG)
Polare Lipide	Monogalactosyldiacylglycerol (MGDG)
	Digalactosyldiacylglycerol (DGDG)
	Phosphatidylglycerol (PG)
	Phosphatidylcholin (PC)
	Phosphatidylethanolamin (PE)
	Phosphatidylinositol (PI)
	Phosphatidylserin (PS)
	Sulfoquinovosyldiacylglycerol

Neutrale Lipide meint bevorzugt Triacylglyceride. Sowohl die neutralen als auch die polaren Lipide können ein breites Spektrum an verschiedenen Fettsäuren enthalten. Beispielhaft jedoch nicht einschränkend seien die in Tabelle 2 aufgeführten Fettsäuren zu nennen.

Tabelle 2: Übersicht über verschiedene Fettsäuren (Auswahl)

[1] Kettenlänge:Anzahl der Doppelbindungen

* nicht natürlicherweise in Pflanzen vorkommend

Nomenklatur[1]	Name
16:0	Palmitinsäure
16:1	Palmitoleinsäure
16:3	Roughaninsäure
18:0	Stearinsäure
18:1	Ölsäure
18:2	Linolsäure
18:3	Linolensäure
γ-18:3	Gamma-Linolensäure*
20:0	Arachidinsäure
22:6	Docosahexanonsäure (DHA) *
20:2	Eicosadienonsäure
20:4	Arachidonsäure (AA) *
20:5	Eicosapentaenosäure (EPA) *
22:1	Erucasäure

Für weitergehende Informationen wird ebenfalls auf das Römpp Chemie Lexikon – CD Version 2.0, Stuttgart/New York: Georg Thieme Verlag 1999 verwiesen.

Erfindungsgemäß eignen sich alle Pflanzen zur Durchführung des erfindungsgemäßen Verfahrens.

"Pflanze" umfasst alle einjährigen und mehrjährige, monokotyledonen und dikotyledonen Pflanzen und schließt beispielhaft jedoch nicht einschränkend solche der Gattungen Cucurbita, Rosa, Vitis, Juglans, Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersicon, Nicotiana, Solarium, Petunia, Digitalis, Majorana, Ciahorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesis, Pelargonium, Panieum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Pisum, Phaseolus, Lolium, Oryza, Zea, Avena, Hordeum, Secale, Triticum, Sorghum, Picea und Populus ein.

Bevorzugt sind Pflanzen nachfolgender Pflanzenfamilien: Amaranthaceae, Asteraceae, Brassicacea, Carophyllaceae, Chenopodiaceae, Compositae, Cruciferae, Cucurbitaceae, Labiatae, Leguminosae, Papilionoideae, Liliaceae, Linaceae, Malvaceae, Rosaceae, Rubiaceae, Saxifragaceae, Scrophulariaceae, Solanacea, Sterculiaceae, Tetragoniacea, Theaceae, Umbelliferae.

Bevorzugte monokotyle Pflanzen sind insbesondere ausgewählt aus den monokotylen Kulturpflanzen, wie zum Beispiel der Familie der Gramineae wie Reis, Mais, Weizen oder andere Getreidearten wie Gerste, Hirse, Roggen, Triticale oder Hafer sowie dem Zuckerrohr sowie alle Arten von Gräsern.
Die Erfindung wird ganz besonders bevorzugt aus dikotyledone pflanzliche Organismen angewendet. Bevorzugte dikotyle Pflanzen sind insbesondere ausgewählt aus den dikotylen Kulturpflanzen, wie zum Beispiel

- Asteraceae wie Sonnenblume, Tagetes oder Calendula und andere mehr,

- Compositae, besonders die Gattung Lactuca, ganz besonders die Art sativa (Salat) und andere mehr,

- Cruciferae, besonders die Gattung Brassica, ganz besonders die Arten napus (Raps), campestris (Rübe), oleracea cv Tastie (Kohl), oleracea cv Snowball Y (Blumenkohl) und oleracea cv Emperor (Broccoli) und weitere Kohlarten; und der Gattung Arabidopsis, ganz besonders die Art thaliana sowie Kresse oder Canola und andere mehr,

- Cucurbitaceae wie Melone, Kürbis oder Zucchini und andere mehr,

- Leguminosae besonders die Gattung Glycine, ganz besonders die Art max (Sojabohne) Soja sowie Alfalfa, Erbse, Bohnengewächsen oder Erdnuss und andere mehr

- Rubiaceae, bevorzugt der Unterklasse Lamiidae wie beispielsweise Coffea arabica oder Coffea liberica (Kaffeestrauch) und andere mehr,

- Solanaceae besonders die Gattung Lycopersicon, ganz besonders die Art esculentum (Tomate) und die Gattung Solanum, ganz besonders die Art tuberosum (Kartoffel) und melongena (Aubergine) sowie Tabak oder Paprika und andere mehr,

- Sterculiaceae, bevorzugt der Unterklasse Dilleniidae wie beispielsweise Theobroma cacao (Kakaostrauch) und andere mehr,

- Theaceae, bevorzugt der Unterklasse Dilleniidae wie beispielsweise Camellia sinensis oder Thea sinensis (Teestrauch) und andere mehr,

- Umbelliferae, besonders die Gattung Daucus (ganz besonders die Art carota (Karrotte)) und Apium (ganz besonders die Art graveolens dulce (Selarie)) und andere mehr; und die Gattung Capsicum, ganz besonders die Art annum (Pfeffer) und andere mehr,

sowie Lein, Soja, Baumwolle, Hanf, Flachs, Gurke, Spinat, Möhre, Zuckerrübe und den verschiedenen Baum-, Nuss- und Weinarten, insbesondere Banane und Kiwi.

Umfasst sind ferner Schmuckpflanzen, Nutz- oder Zierbäume, Blumen, Schnittblumen, Sträucher oder Rasen. Beispielhaft aber nicht einschränkend seien zu nennen Angiospermen, Bryophyten wie zum Beispiel Hepaticae (Leberblümchen) und Musci (Moose); Pteridophyten wie Farne, Schachtelhalm und Lycopoden; Gymnospermen wie Koniferen, Cycaden, Ginkgo und Gnetalen, die Familien der Rosaceae wie Rose, Ericaceae wie Rhododendrons und Azaleen, Euphorbiaceae wie Weihnachtssterne und Kroton, Caryophyllaceae wie Nelken, Solanaceae wie Petunien, Gesneriaceae wie das Usambaraveilchen, Balsaminaceae wie das Springkraut, Orchidaceae wie Orchideen, Iridaceae wie Gladiolen, Iris, Freesie und Krokus, Compositae wie Ringelblume, Geraniaceae wie Geranien, Liliaceae wie der Drachenbaum, Moraceae wie Ficus, Araceae wie Philodendron und andere mehr.

Pflanzliche Organismen im Sinne der Erfindung sind weiterhin weitere photosynthetisch aktive befähigte Organismen, wie zum Beispiel Algen, Cyanobakterien sowie Moose. Bevorzugte Algen sind Grünalgen, wie beispielsweise Algen der Gattung Haematococcus, Phaedactylum tricornatum, Volvox oder Dunaliella. Insbesondere bevorzugt ist Synechocystis.

Am meisten bevorzugt sind Ölpflanzen, d.h. Pflanzen, die bereits natürlicherweise einen hohen Ölgehalt aufweisen und/oder zu industriellen Gewinnung von Ölen verwendet werden. Diese Pflanzen können einen hohen Ölgehaltr und/oder aber eine besondere, industriell interessante Fettsäurezusammensetzung aufweisen. Bevorzugt sind Pflanzen, die einen Lipidanteil von mindestens 1 Gew.% aufweisen. Ölpflanzen umfassen beispielhaft: Bovago oficinalis (Borretsch); Brassica Arten wie *B. campestris, B. napus, B. rapa* (Senf oder Raps); *Cannabis sativa* (Hanf); Curthamus *tinctorius* (Färberdiestel); Cocos *nucifera* (Kokusnuss); Crambe abyssinica (Krambe); Cuphea Arten (Cuphea Arten liefern fettsäuren von mittlerer Kettenlänge insbesondere für industrielle Anwendungen); *Elaeis guinensis* (Afrikanische Ölpalme); *Ekeis oleiferu* (Amerikanische Ölpalme); *Glycine max* (Sojabohne); *Gossypium hirisfum* (Amerikanische Baumwolle); *Gossypium barbadense* (Ägyptische Baumwolle); *Gossypium herbaceum* (Asiatische Baumwolle Cotton); *Helianthus annus* (Sonnenblume); *Linum usitatissimum* (Lein oder Flachs); *Oenethem biennis* (Evening primrose); Ozea *europea* (Olive); Oryza *sativa* (Rice); *Ricinus communis* (Castor); Sesamum *indicum* (Sesam); *Glycine max* (Sojabohne); *Triticum* Arten (Weizen); Zea *maize* (Mais) sowie verschiedene Nussarten wie beispielsweise Walnuss oder Mandel.

Wenn es sich bei den verwendeten Pflanzen um solche zur Gattung der Leguminosen (Hülsenfrüchte) gehörenden Pflanzen handelt, so fällt unter die Erfindung die Expression von Fremdproteinen (Leghemoglobinen, Hemoglobin), d. h. in der Natur nicht-symbiontisch vorkommenden Leghemoglobinen und/oder Hemoglobinen oder die Veränderung der Pflanzen derart, dass sie das natürlich vorkommende Leghemoglobin überexprimieren.

Höchst bevorzugt sind Kartoffeln, Arabidopsis thaliana und Raps.

Vorteilhaft ist es, wenn die vorgenannten Pflanzen ein Leghemoglobin und/oder Hemoglobin ausgewählt aus der Gruppe bestehend aus Leghemoglobin und/oder Hemoglobin aus den Pflanzen Lupinus luteus (LGB1_LUPLU, LGB2_LUPLU), Glycine max (LGBA_SOYBN, LGB2_SOYBN, LGB3_SOYBN), Medicago sativa (LGB1-4_MEDSA), Medicago trunculata (LGB1_MEDTR), Phaseolus vulgaris (LGB1_PHAVU, LGB2_PHAVU), Vicia faba (LGB1_VICFA, LGB2_VICFA), Pisum sativum (LGB1_PEA, LGB2_PEA), Vigna unguiculata (LGB1_VIGUN), Lotus japonicus (LGB_LOTJA), Psophocarpus tetragonolobus (LGB_PSOTE), Sesbania rostrata (LGB1_SESRO), Casuarina glauca (HBPA_CASGL) und Canvalaria lineata (HBP_CANLI) Physcomitrella patens (HBL0_PHYPA), Arabidopsis thaliana (HBL1_ARATH, HBL2_ARATH), Gossypium hirsutum (HBL1_GOSHI, HBL2_GOSHI), Medicago sativa (HBL1_MEDSA), Oryza sativa (HBL1_ORYSA, HBL2_ORYSA, HBL3_ORYSA, HBL4_ORYSA), Brassica napus (HBL2_BRANA), Lycopersicon esculentum (HBL2_LYCES), Hordeum vulgare (HBL_HORVU), Zea mays (HBL_MAIZE), Trema tomentosa (HBL_TRETO), Casuarina glauca (HBP1_CASGL, HBP2_CASGL, HBPA_CASGL), Parasponia rigida (HBPL_PARAD) exprimieren. In Klammern sind die jeweiligen Swiss-Prot Datenbankeinträge vermerkt.

Besonders vorteilhaft sind Pflanzen, die die Sequenz Nr. 1 codierend für Leghemoglobin und/oder die Sequenzen Nr. 3 und/oder 5 codierend für Hemoglobin aufweisen. Bevorzugterweise wird Leghemoglobin und/oder Hemoglobin aus *Lotus japonicus* eingesetzt. Dann produzieren die transformierten Pflanzen Speicherstoffe in erhöhtem Ausmaß.

Unter die Erfindung fallen auch Pflanzen, die zur Produktion von Speicherstoffen ein Leghemoglobin, gemäß Sequenz Nr. 1 und/oder ein Hemoglobin gemäß Sequenzen Nr. 3 und 5 exprimieren.

In einer bevorzugten Variante der Erfindung handelt es sich um Pflanzen, die das Leghemoglobin und/oder Hemoglobin speicherorganspezifisch exprimieren.

Die genannten Pflanzen lagern die Speicherstoffe in der Regel in bestimmten Organen ab. Dies sind z. B. Zwiebeln, Knollen, Samen, Körner, Nüsse, Blätter usw. Unter Speicherorgane im Sinne der Erfindung sind auch Früchte zu verstehen. Früchte sind die Sammelbezeichnung für die den Samen als Nährgewebe umhüllenden Organe der Pflanzen. Dabei ist nicht nur an die eßbaren Früchte, insbesondere an das Obst, aber auch an Hülsenfrüchte, Getreide, Nüsse, Gewürze zu denken, sondern auch an offiziell genutzte Drogen (s. Fructus, Semen). Natürlich kann auch eine Speicherung der Speicherstoffe in der gesamten Pflanze stattfinden.

Bevorzugterweise wird das Leghemoglobin und/oder Hemoglobin knollenspezifisch oder samenspezifisch exprimiert.

Als Pflanzen kommen alle zuvor genannten in Betracht. Bevorzugt werden Kartoffeln, Arabidopsis thaliana, Raps, Sojabohnen, Erdnüsse, Mais, Maniok, Yams, Reis, Sonnenblumen, Roggen, Gerste, Hopfen, Hafer, Hart und Weichweizen, Lupinen, Erbsen, Klee, Rüben, Kohl, Reben usw. wie sie z. B. aus der Verordnung über das Artenverzeichnis zum Saatgutverkehrsgesetz (Blatt für PMZ 1986 S. 3, zuletzt geändert Blatt für PMZ 2002 S. 68) entnehmbar sind, eingesetzt.

Es ist besonders bevorzugt, wenn es sich um knollenproduzierende Pflanzen, insbesondere Kartoffel-Pflanzen oder um samenproduzierende Pflanzen, insbesondere Arabidopsis thaliana, Raps oder Soja handelt.

Die gewebespezifische Expression kann z. B. durch Verwendung eines gewebespezifischen Promotors erreicht werden. Eine solche gewebespezifische Expression ist beispielsweise bekannt aus der US 6,372,961 B1 Spalte 11, Zeilen 44 ff.

Gegenstand der Erfindung ist ferner eine Nukleotidsequenz gemäß Sequenz Nr. 1 codierend für Leghemoglobin zur Verwendung in einer Pflanze und eine entsprechende Genstruktur oder Vektor sowie deren Verwendung zur Transformierung einer Pflanze mit der Erfindung.

Insbesondere fällt unter die Erfindung auch die Verwendung einer Leghemoglobin codierenden Nukleotidsequenz, die mit der Sequenz Nr. 1 zu ca. 70 %, bevorzugt 80%, besonders bevorzugt 85%, 90%, insbesondere 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% oder 99% identisch ist.

Gegenstand der Erfindung sind außerdem Nukleotidsequenzen gemäß Nr. 3 und 5 kodierend für Hemoglobin zur Verwendung in einer Pflanze und eine entsprechende Genstruktur oder -vektor sowie deren Verwendung zur Transformierung einer Pflanze. Insbesondere fallen unter die Erfindung auch die Verwendung von Hemoglobin kodierenden Nukleotidsequenzen, die mit den Sequenzen 3 und 5 zu ca. 70%, bevorzugt 80%, besonders bevorzugt 85%, 90%, insbesondere 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% oder 99% identisch sind.

Unter Nukleotidsequenz werden alle Nukleotidsequenzen verstanden, die (i) exakt den dargestellten Sequenzen entsprechen; oder (ii) mindestens eine Nukleotidsequenz umfaßen, die innerhalb des Bereichs der Degeneration des genetischen Codes den dargestellten Sequenzen entspricht; oder (iii) mindestens eine Nukleotidsequenz umfasst, die mit einer zur Nukleotidsequenz (i) oder (ii) komplemetären Nukleotidsequenz hybridisiert, und gegebenenfalls (iiii) funktionsneutrale Sinnmutationen in (i) umfasst. Dabei bedeutet der Begriff "funktionsneutrale Sinnmutationen" den Austausch chemisch ähnlicher Aminosäuren, wie z. B. Glycin durch Alanin oder Serin durch Threonin.

Erfindungsgemäß sind auch die den kodierenden Bereichen (Strukturgenen) vorausgehenden (5`-oder upstream) und/oder nachfolgenden (3`-oder downstream) Sequenzbereiche eingeschlossen. Insbesondere sind hierin Sequenzbereiche mit regulatorischer Funktion inbegriffen. Sie können die Transkription, die RNA-Stabilität oder das RNA Processing sowie die Translation beeinflussen. Beispiele für regulatorische Sequenzen sind u. a. Promotoren, Enhancer, Operatoren, Teminatoren oder Translationsverstärker.

Unter die jeweiligen Proteine (Leghemoglobine und/oder Hemoglobine) fallen auch Isoformen, die als Proteine mit gleicher oder vergleichbarer Wirkung verstanden werden, die jedoch eine unterschiedliche Primärstruktur aufweisen.

Unter modifizierten Formen sind erfindungsgemäß Proteine zu verstehen, bei denen Änderungen in der Sequenz, beispielsweise am N- und/oder C-Teminus des Polypeptids oder im Bereich konservierter Aminosäuren vorliegen, ohne jedoch die Funktion des Proteins zu beeinträchtigen. Diese

Veränderungen können in Form von Aminosäureaustauschen nach bekannten Methoden vorgenommen werden.

Die Erfindung wird unter Bezugnahme auf die folgenden Versuche beispielhaft beschrieben.

Beispiele

1. Modellorganismen

Als Modellorganismen für die Experimente wurden Kartoffel und Arabidopsis thaliana eingesetzt. Beide Pflanzenarten repräsentieren ein Mitglied der höheren Pflanzen (Samenpflanzen). Sie können aufgrund des hohen Grades an Homologie ihrer DNA-Sequenzen bzw. Polypeptidsequenzen als Modellpflanzen für andere Pflanzenarten eingesetzt werden.

2. Allgemeine Verfahren

a) Anzucht von Kartoffel oder Arabidopsis Pflanzen

Arabidopsis Pflanzen wurden entweder auf Murashige-Skoog Medium mit 0,5 % Saccharose (Ogas et al., 1997 Science 277:91-94) oder auf Erde gezogen (Focks & Benning, 1998 Plant Physiology 118:91-101). Um einheitliche Keimungs- und Blühzeiten zu erreichen, wurden die Samen nach Ausplattieren bzw. Ausstreuen auf Erde zwei Tage bei 4 °C stratifiziert. Nach der Blüte wurden die Schoten markiert. Entsprechend der Markierungen wurden dann Schoten mit einem Alter von 6-20 Tagen nach der Blüte geerntet. Kartoffelpflanzen wurden nach Dietze et al., 1995 (Gene Transfer to Plants, ed. Potrykus und Spangenberg, Springer Lab Manual, Berlin, Heidelberg, 24-29) angezogen.

b) Isolierung von totalRNA und poly-A+ RNA aus Pflanzen

Für die Herstellung von Expressionskonstrukten wird RNA bzw. polyA+ RNA isoliert. RNA wurde aus Schoten von Arabidopis oder aus Knollen von Kartoffel bzw. Wurzeln von Lotus japonicus nach folgender Vorschrift isoliert:

Pflanzenmaterial im Alter von 6 - 40 Tage wurde geerntet und in flüssigem Stickstoff schockgefroren. Das Material wurde vor der weiteren Verwendung bei -80 °C gelagert. 75 mg des Materials wurden im gekühlten Mörser zu einem feinen Pulver gemahlen und mit 200 µL des Lysis-Puffers aus dem Ambion RNAqueos-Kit versetzt. Die Isolierung der totalen RNA wurde dann nach Herstellerangaben durchgeführt. Die RNA wurde mit 50 µL Elutionspuffer (Ambion) eluiert und die Konzentration durch Absorption einer 1.100 verdünnten Lösung am Photometer (Eppendorf) bei 260 nm bestimmt. 40ug/ml RNA entspricht dabei einer Absorption von 1. Die RNA-Lösungen wurden mit RNAse freiem Wasser auf eine Konzentration von 1 µg/µL eingestellt. Die Konzentrationen wurden durch Agarosegelelektrophorese überprüft.

Zur Isolierung von polyA+ RNA wurde oligo(dT)-Zellulose von Amersham Pharmacia nach Herstellerangaben verwendet. RNA bzw. polyA+ RNA wurde bei -70 °C gelagert.

3. Konstruktion der cDNA-Bank

Zur Konstruktion der cDNA-Bank aus Lotus japonicus- bzw. Arabidopsis thaliana-RNA wurde die Erststrangsynthese unter Verwendung von Reverser Transkriptase aus Maus-Leukämie-Virus (Clontech) und Oligo-d(T)-Primern, die Zweitstrangsynthese durch Inkubation mit DNA-Polymerase I, Klenow-Enzym und RNAse H-Spaltung bei 12 °C (2 Std.), 16 °C (1 Std.) und 22 °C (1 Std.) erzielt. Die Reaktion wurde durch Inkubation bei 65 °C (10 min)

gestoppt und anschließend auf Eis überführt. Doppelsträngige DNA-Moleküle wurde mit T4-DNA-Polymerase (Roche, Mannheim) bei 37 °C (30 min) mit glatten Enden versehen. Die Nukleotide wurden durch Phenol/Chloroform-Extraktion und Sephadex-G50-Zentrifugiersäulen entfernt. EcoRI/NotI-Adapter (Pharmacia, Freiburg, Deutschland) wurden mittels T4-DNA-Ligase (Roche, 12°C, über Nacht) an die cDNA-Enden ligiert, mit NotI nachgeschnitten und durch Inkubation mit Polynukleotidkinase (Roche, 37 °C, 30 min) phosphoryliert. Dieses Gemisch wurde der Trennung auf einem Low-Melting-Agarose-Gel unterworfen. DNA-Moleküle über 200 Basenpaaren wurden aus dem Gel eluiert, Phenol-extrahiert, auf Elutip-D-Säulen (Schleicher und Schüll, Dassel, Deutschland) konzentriert und in den Klonierungsvektor pSPORT1 (Invitrogen, Karlsruhe) ligiert, wobei Material des Herstellers verwendet und seine Anweisungen befolgt wurden.

4. DNA-Sequenzierung und Computeranalyse

cDNA-Banken, wie unter Punkt 2 beschrieben, wurden zur DNA-Sequenzierung nach Standardverfahren, insbesondere durch das Kettenterminationsverfahren unter Verwendung des ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction-Kit (Perkin-Elmer, Weiterstadt, Deutschland), verwendet. Die Sequenzierung zufälliger, vereinzelter Klone wurde anschließend an die präparative Plasmidgewinnung aus cDNA-Banken über in vivo-Massenexcision und Retransformation von DH5α auf Agarplatten durchgeführt (Einzelheiten zu Material und Protokoll von Stratagene, Amsterdam, Niederlande). Plasmid-DNA wurde aus über Nacht gezüchteten E. coli-Kulturen, die in Luria-Brühe mit Ampicillin (siehe Sambrook et al. (1989) (Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6)) gezüchtet worden waren, an einem Qiagen-DNA-Präparations-Roboter (Qiagen, Hilden) nach den Protokollen des Herstellers präpariert.

Die Sequenzen wurden unter Verwendung des Standard-Softwarepakets EST-MAX, das kommerziell von Bio-Max (München, Deutschland) geliefert wird, prozessiert und annotiert. Durch Nutzung von Vergleichsalgorithmen und unter Verwendung einer Suchsequenz wurde mithilfe des BLAST-Programms nach homologen Genen gesucht (Altschul et al. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402).

5. Herstellung der Expressionskonstrukte für die knollenspezifische Expression in Kartoffel bzw. die samenspezifische Expression in Arabidopsis:

a) Knollenspezifische Expression:

Das Leghemoglobin-Gen Ljlb wurde aus dem Vektor pSportI mit NotI und KpnI geschnitten, wobei die NotI Schnittstelle durch Inkubation mit Klenow-Enzym aufgefüllt wurde. Für die Ligation in den Vektor pART33 wurde dieser mit BamHI verdaut, dann die Schnittstelle durch Inkubation mit Klenow-Enzym aufgefüllt und anschliessend mit KpnI verdaut. pART33 enthält bereits den B33-Promoter (1459 bp; Liu XJ, Prat S, Willmitzer L, Frommer WB (1990) cis regulatory elements directing tuber-specific and sucrose-inducible expression of a chimeric class I patatin promoter/GUS-gene fusion. Mol Gen Genet. 223(3):401-6) sowie den OCS-Terminator (766 bp) für die knollenspezifische Expression. Das gesamte Konstrukt wurde mit NotI aus pART33 geschnitten und in den Vektor für die Pflanzentransformation pART27 einkloniert.

b) Samenspezifische Expression:

Das Leghemoglobin-Gen Ljlb wurde aus dem Vektor pART33 in den Vektor pBINUSP umkloniert. Zu diesem Zweck wurde das Konstrukt pART33-LegHb zunächst mit Asp718I verdaut und die Überhänge mit Klenow-Fragment

aufgefüllt. Anschließend wurde das linearisierte Konstrukt mit XbaI-verdaut und das so gewonnene Gen über Nacht bei 4 °C in den XbaI/SmaI-geschnittenen pBINUSP ligiert.

Das Hemoglobin-Gen1 AtHb1 wurde aus dem Vektor pART33 in den Vektor pBINUSP umkloniert. Zu diesem Zweck wurde das Konstrukt pART33-AtHb1 zunächst mit Asp718I verdaut und die Überhänge mit Klenow-Fragment aufgefüllt. Anschließend wurde das linearisierte Konstrukt mit XbaI-verdaut und das so gewonnene Gen über Nacht bei 4 °C in den XbaI/SmaI-geschnittenen pBINUSP ligiert.

Das Hemoglobin-Gen2 AtHb2 wurde aus dem Vektor pART33 in den Vektor pBINUSP umkloniert. Zu diesem Zweck wurde das Konstrukt pART33-AtHb2 zunächst mit Asp718I verdaut und die Überhänge mit Klenow-Fragment aufgefüllt. Anschließend wurde das linearisierte Konstrukt mit XbaI-verdaut und das so gewonnene Gen über Nacht bei 4 °C in den XbaI/SmaI-geschnittenen pBINUSP ligiert.

6. Plasmide für die Pflanzentransformation

Zur Pflanzentransformation können binäre Vektoren, wie pBIN verwendet werden (Bevan, M., Nucleic Acids Res. 12 (1984), 8711–8721). Die Konstruktion der binären Vektoren kann durch Ligation der cDNA in Sense- oder Antisense-Orientierung in T-DNA erfolgen. 5' der cDNA aktiviert ein Pflanzenpromotor die Transkription der cDNA. Eine Polyadenylierungssequenz befindet sich 3' von der cDNA.

Die gewebespezifische Expression läßt sich unter Verwendung eines gewebespezifischen Promotors erzielen. Beispielsweise kann die samenspezifische Expression erreicht werden, indem der Napin- oder der LeB4- oder der USP-Promotor 5' der cDNA einkloniert wird. Auch jedes andere samenspezifische Promotorelement kann verwendet werden. Zur

konstitutiven Expression in der ganzen Pflanzen läßt sich der CaMV-35S-Promotor verwenden.

7. Transformation von Agrobacterium und Pflanzentransformation

Die Agrobacterium-vermittelte Pflanzentransformation kann zum Beispiel unter Verwendung des GV3101- (pMP90-) (Koncz und Schell, Mol. Gen. Genet. 204 (1986) 383-396) oder LBA4404- (Clontech) Agrobacterium tumefaciens-Stamms durchgeführt werden. Die Transformation des Bakteriums kann durch dem Fachmann bekannte Standard-Transformationstechniken durchgeführt werden (Deblaere et al., Nucl. Acids. Tes. 13 (1984), 4777-4788).

Die Agrobacterium-vermittelte Pflanzentransformation selbst kann unter Verwendung von dem Fachmann ebenfalls bekannte Standard-Transformations- und Regenerationstechniken durchgeführt werden (Gelvin, Stanton B., Schilperoort, Robert A., Plant Molecular Biology Manual, 2. Aufl., Dordrecht: Kluwer Academic Publ., 1995, in Sect., Ringbuc Zentrale Signatur: BT11-P ISBN 0-7923-2731-4; Glick, Bernard R., Thompson, John E., Methods in Plant Molecular Biology and Biotechnology, Boca Raton: CRC Press, 1993, 360 S., ISBN 0-8493-5164-2).

Die Transformation mittels Agrobacterium von Arabisopsis thaliana wurde nach der Methode von Bechthold et al., 1993 (C.R. Acad. Sci. Ser. III Sci. Vie., 316, 1194-1199) durchgeführt. Die Methode wurde dahingegen modifiziert, dass auf die Vakuuminfiltration verzichtet wurde.

Für die Transformation wurden *Arabidopsis thaliana* Col0 Samen auf befeuchteter Erde ausgesät, zwei Tage bei 4 °C stratifiziert und 4-6 Wochen unter Kurztagbedingungen (8 h Licht, 21 °C) angezogen. Die Pflanzen wurden dann zur Blühinduktion in Langtagbedingungen (16 h Licht, 21 °C)

überführt. Nach etwa 10 Tagen ist die Infloreszenz (Blütenstand) gross genug für das Eintauchen. Die Influoreszenz wird in eine *Agrobacterium* Suspension mit ½ MS Lösung (Murashige und Skoog, s.u.) pH 5,7, 5 % Saccharose, 44 µM Benzylaminopurin (Sigma) und 0,03 % Silwet L-77 (Lehle Seeds, USA) getaucht. Die optische Dichte der *Agrobacterium* Suspension bei 600 nm sollte 0,8 betragen. Die Bakterien werden zuvor im YEB Medium (0,5 % Bactotrypton, 0,5 % Bactopeptone, 0,1 % Hefeextrakt, 0,5 % Saccharose und 2mM $MgCl_2$) angezogen.

Nach Eintauchen der Pflanzen wurden diese für etwa drei Wochen weiter befeuchtet. Dann wurde die Bewässerung eingestellt und die Pflanzen trockneten für die Samengewinnung ab. Die erhaltenen Samen wurden auf Platten mit ½ MS Lösung, pH 5,7, 50 µg Kanamycin, 250 µg Timenten und 0,8 % Agar ausgelegt. Resistente Keimlinge wurden vereinzelt und auf Erde pikiert.

Die Transformation von Kartoffel (Solanum tuberosum L.) erfolgte nach der Methode von Dietze et al. in Gene transfer to plants (eds. I. Potrykus and G. Spangenberg, Springer Lab Manual, 1995, S. 24-29). Als Ausgangsmaterial wurde die Sorte Desiree verwendet.
Zur Transformation wurden fünf bis sechs Blätter einer sterilen Sprossen-Kultur der Kartoffelvarietät Desiree vorsichtig in 10 ml einer 2 MS Lösung (Standard MS Medium nach Murashige and Skoog (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.*, 15, 473-497.) mit zusätzlich 2 % Saccharose equilibriert. Aus den Blättern werden vertikal zur Mittelrippe 1 - 2 mm lange Stücke herausgeschnitten und in frisches 2 MS Medium gegeben. Zu dem frischen Medium werden 50 µl einer Agrobacterium-Lösung zugegeben, die mit dem T-Plasmid pART27-JpLeg transformiert wurden. Die Agrobacterium-Lösung wird über die

Blattstücke verteilt und die Platten werden für zwei Tage im Dunkeln bei 22 - 24 °C inkubiert. Nach zwei Tagen werden die Blattstücke auf CIM Medium (MS Medium mit 1,5 % Glucose, 5 mg/l NAA, 0,1 mg/l BAP, 250 mg/l Claforin und 50 mg/l Kanamycin oder Hygromycin) umgesetzt und 7 Tage inkubiert. Dann werden die Blattstücke auf SIM Medium (MS Medium mit 2 mg/l Zeatin, 0,02 mg/l NAA, 0,02 mg/l GA3, 250 mg/l Claforin und 50 mg/l Kanamycin oder Hygromycin) umgesetzt. Nach 1 - 2 Wochen können dann die entstandenen transgenen Sprosse (1 - 1,5 cm lang) abgeschnitten und auf RIM Medium (MS Medium mit 250 mg/l Claforin) umgesetzt werden. Nach 3 - 4 Wochen haben die Sprossen Blätter und Wurzel ausgebildet und die Pflanzen können auf Erde pikiert werden.

Entsprechend dieser Methode können 2 - 3 unabhängige transgene Linien pro eingesetztes Blatt erhalten werden.

Die alternativ mögliche Pflanzentransformation unter Verwendung von Teilchenbeschuß, Polyethylenglycol-vermittelter DNA-Aufnahme oder über die Siliziumcarbonatfaser-Technik ist beispielsweise beschrieben von Freeling und Walbot "The maize handbook" (1993) ISBN 3-540-97826-7, Springer Verlag New York).

8. Untersuchung der Expression des Leghemoglobins und Hemoglobins in der transformierten Pflanze

Die Aktivität des Leghemoglobins in den transformierten Pflanzen wurde auf der Transkriptionsebene gemessen.

Ein geeignetes Verfahren zur Bestimmung der Menge an Transkription des Gens (ein Hinweis auf die Menge an RNA, die für die Translation des Genproduktes zur Verfügung steht) ist die Durchführung eines Northern-Blots

wie unten ausgeführt (als Bezugsstelle siehe Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York, oder den oben erwähnten Beispielteil), wobei ein Primer, der so gestaltet ist, daß er an das Leghemoglobin bindet, mit einer nachweisbaren Markierung (gewöhnlich radioaktiv oder chemilumineszent) markiert wird, so daß, wenn die Gesamt-RNA einer Kultur des Organismus extrahiert, auf einem Gel aufgetrennt, auf eine stabile Matrix transferiert und mit dieser Sonde inkubiert wird, die Bindung und das Ausmaß der Bindung der Sonde das Vorliegen und auch die Menge der mRNA für dieses Gen (Leghemoglobin bzw. Hemoglobin) anzeigt. Diese Information zeigt den Grad der Transkription des transformierten Gens an. Zelluläre Gesamt-RNA kann aus Zellen, Geweben oder Organen mit mehreren Verfahren, die alle im Fachgebiet bekannt sind, wie zum Beispiel das von Bormann, E. R., et al. (1992) Mol. Microbiol. 6:317-326 beschriebene, präpariert werden.

Northern-Hybridisierung:

Für die RNA-Hybridisierung wurden 20 µg Gesamt-RNA oder 1 µg poly(A)$^+$-RNA mittels Gelelektrophorese in Agarosegelen mit einer Stärke von 1,25 % unter Verwendung von Formaldehyd, wie beschrieben in Amasino (1986, Anal. Biochem. 152, 304) aufgetrennt, mittels Kapillaranziehung unter Verwendung von 10 x SSC auf positiv geladene Nylonmembranen (Hybond N+, Amersham, Braunschweig) übertragen, mittels UV-Licht immobilisiert und 3 Stunden bei 68 °C unter Verwendung von Hybridisierungspuffer (10 % Dextransulfat Gew./Vol., 1 M NaCl, 1 % SDS, 100 mg Heringssperma-DNA) vorhybridisiert. Die Markierung der DNA-Sonde mit dem Highprime DNA Labeling-Kit (Roche, Mannheim, Deutschland) erfolgte während der Vorhybridisierung unter Verwendung von alpha-^{32}P-dCTP (Amersham Pharmacia, Braunschweig, Deutschland). Die Hybridisierung wurde nach Zugabe der markierten DNA-Sonde im gleichen Puffer bei 68 °C über Nacht

durchgeführt. Die Waschschritte wurden zweimal für 15 min unter Verwendung von 2 X SSC und zweimal für 30 min unter Verwendung von 1 X SSC, 1 % SDS, bei 68 °C durchgeführt. Die Exposition der verschlossenen Filter wurde bei -70 °C für einen Zeitraum von 1 bis 14 Tagen durchgeführt.

9. Analyse der Auswirkung des rekombinanten Leghemoglobins auf die Produktion des gewünschten Produktes

Die Auswirkung der genetischen Modifikation in Pflanzen, Pilzen, Algen, Ciliaten oder auf die Produktion einer gewünschten Verbindung (wie einer Fettsäure) kann bestimmt werden, indem die modifizierten Mikroorganismen oder die modifizierte Pflanze unter geeigneten Bedingungen (wie den vorstehend beschriebenen) gezüchtet werden und das Medium und/oder die zellulären Komponenten auf die erhöhte Produktion des gewünschten Produktes (d.h. von Lipiden oder Kohlenhydrate) untersucht wird. Diese Analysetechniken sind dem Fachmann bekannt und umfassen Spektroskopie, Dünnschichtchromatographie, Färbeverfahren verschiedener Art, enzymatische und mikrobiologische Verfahren sowie analytische Chromatographie, wie Hochleistungs-Flüssigkeitschromatographie (siehe beispielsweise Ullman, Encyclopedia of Industrial Chemistry, Bd. A2, S. 89-90 und S. 443-613, VCH: Weinheim (1985); Fallon, A., et al., (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, Bd. 17; Rehm et al. (1993) Biotechnology, Bd. 3, Kapitel III: "Product recovery and purification", S. 469-714, VCH: Weinheim; Belter, P.A., et al. (1988) Bioseparations: downstream processing for Biotechnology, John Wiley and Sons; Kennedy, J.F., und Cabral, J.M.S. (1992) Recovery processes for biological Materials, John Wiley and Sons; Shaeiwitz, J.A., und Henry, J.D. (1988) Biochemical Separations, in: Ullmann's Encyclopedia of Industrial Chemistry, Bd. B3;

Kapitel 11, S. 1-27, VCH: Weinheim; und Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications).

Neben den oben erwähnten Verfahren werden Pflanzenlipide aus Pflanzenmaterial wie von Cahoon et al. (1999) Proc. Natl. Acad. Sci. USA 96 (22):12935-12940, und Browse et al. (1986) Analytic Biochemistry 152:141-145, beschrieben extrahiert. Die qualitative und quantitative Lipid- oder Fettsäureanalyse ist beschrieben bei Christie, William W., Advances in Lipid Methodology, Ayr/Scotland: Oily Press (Oily Press Lipid Library; 2); Christie, William W., Gas Chromatography and Lipids. A Practical Guide - Ayr, Scotland: Oily Press, 1989, Repr. 1992, IX, 307 S. (Oily Press Lipid Library; 1); "Progress in Lipid Research, Oxford: Pergamon Press, 1 (1952) - 16 (1977) u.d.T.: Progress in the Chemistry of Fats and Other Lipids CODEN.

Zusätzlich zur Messung des Endproduktes der Fermentation ist es auch möglich, andere Komponenten der Stoffwechselwege zu analysieren, die zur Produktion der gewünschten Verbindung verwendet werden, wie Zwischen- und Nebenprodukte, um die Gesamteffizienz der Produktion der Verbindung zu bestimmen. Die Analyseverfahren umfassen Messungen der Nährstoffmengen im Medium (z.B. Zucker, Kohlenwasserstoffe, Stickstoffquellen, Phosphat und andere Ionen), Messungen der Biomassezusammensetzung und des Wachstums, Analyse der Produktion üblicher Metabolite von Biosynthesewegen und Messungen von Gasen, die während der Fermentation erzeugt werden. Standardverfahren für diese Messungen sind in Applied Microbial Physiology; A Practical Approach, P.M. Rhodes und P.F. Stanbury, Hrsgb., IRL Press, S. 103-129; 131-163 und 165-192 (ISBN: 0199635773) und darin angegebenen Literaturstellen beschrieben.

Ein Beispiel ist die Analyse von Fettsäuren (Abkürzungen: FAME, Fettsäuremethylester; GC-MS, Gas-Flüssigkeitschromatographie-Massenspektrometrie; TAG, Triacylglycerin; TLC, Dünnschichtchromatographie).

Der unzweideutige Nachweis für das Vorliegen von Fettsäureprodukten kann mittels Analyse rekombinanter Organismen nach Standard-Analyseverfahren erhalten werden: GC, GC-MS oder TLC, wie verschiedentlich beschrieben von Christie und den Literaturstellen darin (1997, in: Advances on Lipid Methodology, Vierte Aufl.: Christie, Oily Press, Dundee, 119-169; 1998, Gaschromatographie-Massenspektrometrie-Verfahren, Lipide 33:343-353).

Das zu analysierende Material kann durch Ultraschallbehandlung, Mahlen in der Glasmühle, flüssigen Stickstoff und Mahlen oder über andere anwendbare Verfahren aufgebrochen werden. Das Material muss nach dem Aufbrechen zentrifugiert werden. Das Sediment wird in Aqua dest. resuspendiert, 10 min bei 100 °C erhitzt, auf Eis abgekühlt und erneut zentrifugiert, gefolgt von Extraktion in 0,5 M Schwefelsäure in Methanol mit 2 % Dimethoxypropan für 1 Std. bei 90 °C, was zu hydrolysierten Öl- und Lipidverbindungen führt, die transmethylierte Lipide ergeben. Diese Fettsäuremethylester werden in Petrolether extrahiert und schließlich einer GC-Analyse unter Verwendung einer Kapillarsäule (Chrompack, WCOT Fused Silica, CP-Wax-52 CB, 25 mikrom, 0,32 mm) bei einem Temperaturgradienten zwischen 170 °C und 240 °C für 20 min und 5 min bei 240 °C unterworfen. Die Identität der erhaltenen Fettsäuremethylester muss unter Verwendung von Standards, die aus kommerziellen Quellen erhältlich sind (d.h. Sigma), definiert werden.

10. Analyse der Expression von Lotus-Leghemoglobin, Arabidopsis-Hemoglobin1, Arabidopsis-Hemoglobin2 in reifenden Samen von Arabidopsis.

Zur Überprüfung der Expression der Transgene Hemoglobin1 bzw. Hemoglobin 2 aus Arabidopsis thaliana (AtHb1 bzw. AtHb2) sowie Leghemoglobin aus Lotus japonicus (LjLegHb) wurden Northern Blot-Analysen durchgeführt. Hierzu wurde gesamt RNA aus reifenden Samen transgener und Wildtyp-Pflanzen isoliert, elektrophoretisch aufgetrennt und die Transkripte mit den entsprechenden Digoxigenin-markierten Sonden detektiert. Für jede Linie wurden 2-4 unabhängige Analysen durchgeführt. Abbildung 1 zeigt beispielhaft die Ergebnisse der Expressionsanalysen mit reifenden T2 Samen transgener Arabidopsis-Linien, die entweder Lotus-Hemoglobin (LjLegHb), Arabidopsis-Hemoglobin1 (AtHb1) oder Arabidosis-Hemoglobin2 (AtHb2) exprimieren. In fast allen analysierten Linien wird das transformierte Gen transkribiert, da die entsprechende mRNA nachgewiesen werden kann. In den Kontrollen kann hingegen kein Transkript nachgewiesen werden.

11. Analyse des Ölgehalts von transformierten Lotus-Legehemoglobin und Arabidopsis-Hemoglobine exprimierenden Arabidopsis-Pflanzen

Die Bestimmung der Ölgehalte in den Samen transgener Arabidopsis-Pflanzen, die Lotus-Leghemoglobin, Arabidopsis-Hemoglobin1 oder Arabidopsis Hemoglobin2 exprimieren, erfolgt indirekt über die Quantifizierung der Gesamtfettsäuren. Hierzu wurde folgendes Protokoll angewendet:
Die Extraktion der Lipide aus den Samen erfolgt nach der Methode von Blight & Dyer, 1959 Can. J. Biochem. Physiol. 37:911. Für jede Messung werden

hierzu 5-10 Samen in 1,2 ml Qiagen-Microtubes (Qiagen, Hilden) überführt, in einer Retsch Schwingmühle (MM300, Firma Retsch (Haan) pulverisiert und die Gesamtlipide durch Zugabe von 500 µl Chloroform/Methanol (2:1, enthält Pentadecansäure (C15) als internen Standard) extrahiert. Nach Zugabe von 500 µl 50 mM Kaliumphosphatpuffer pH 7,5 erfolgt die Phasentrennung. Die organische Phase wird in ein Pyrrex-Röhrchen überführt und bis zur Trockene eingengt. Anschließend erfolgt die Derivatisierung der Gesamtfettsäuren zu Fettsäuremethylestern nach der Methode von Benning & Sommerville, 1992 J. Bacteriol. 174: 6479-6487 durch Zugabe von 1N H_2SO_4 in Methanol und 2 % (v/v) Dimethoxypropane für 1 h bei 80 °C. Diese Fettsäuremethylester werden in Petrolether extrahiert und schließlich gaschromatographisch unter Verwendung einer Kapillarsäule (Chrompack, WCOT Fused Silica, CP-Wax-52, 25 mikom, 0,32 mm) analysiert. Die Quantifizierung der Gesamtfettsäuren erfolgt durch Vergleich der Signalflächen der Fettsäuremethylester mit der Signalfläche des internen Standards bekannter Konzentration. Ein weiterer Indikator für den Ölgehalt in Arabidopsis thaliana ist die Menge der Fettsäure Eicosensäure ($20:1^{\Delta 11}$), welche fast ausschließlich in den Speicherlipiden vorkommt.

Nach der Transformation von Arabidopsis thaliana werden transgene Linien über Antibiotikaresistenz selektioniert. Hierzu wurden T1 Samen transformierter Arabidopsis-Pflanzen aus Hygromycin-haltigen Selektionsplatten ausgekeimt. Für die quantitative Bestimmung der Ölgehalte wurden T2 Samen von jeweils 4 unabhängigen transgenen Linien analysiert. Für die einzelnen Linien wurden hierbei 5-9 unabhängige Einzelmessungen durchgeführt und die entsprechenden Mittelwerte gebildet.

Die Ergebnisse der Bestimmung der Ölgehalte in reifen T2 Samen transgener Pflanzen, die Lotus-Leghemoglobin exprimieren, sind beispielhaft in der Tabelle 3 und in der Abbildung 2 gezeigt. Gemäß der Erfindung führt die

Expression des Lotus-Leghemoglobins in den Linien 4 und 5 zu einer signifikanten Erhöhung des Ölgehaltes um 17,14 % bzw. 17,54 %.

Die Ergebnisse der Bestimmung der Ölgehalte in reifen T2 Samen transgener Pflanzen, die Arabidopis-Hemoglobin1 exprimieren, sind beispielhaft in der Tabelle 4 und in der Abbildung 3 gezeigt. Gemäß der Erfindung führt die Expression des Arabidopsis-Hemoglobins1 in den Linien 35, 13 und 39 zu einer signifikanten Erhöhung des Ölgehaltes um 9,78 %, 12,22 % bzw. 51,16 %.

Die Ergebnisse der Bestimmung der Ölgehalte in reifen T2 Samen transgener Pflanzen, die Arabidopis-Hemoglobin2 exprimieren, sind beispielhaft in der Tabelle 5 und in der Abbildung 4 gezeigt. Gemäß der Erfindung führt die Expression des Arabidopsis-Hemoglobins 2 in den Linien 2, 10 und 11 zu einer signifikanten Erhöhung des Ölgehaltes um 16,25 %, 21,86 % bzw. 23,80 %.

12. Bestimmung des Stärkegehalts in transformierten Lotus-Leghemoglobin exprimierenden Kartoffelpflanzen

Zur Bestimmung des Stärkegehaltes von Knollen der Kartoffelpflanzen wird die Dichtemessung nach Schéele C. von, Svensson, G. and Rasmusson J., Die Bestimmung des Stärkegehalts und der Trockensubstanz der Kartoffel mit Hilfe des spezifischen Gewichts. Landw. Vers Sta. 127: 67-96, 1937 eingesetzt. Anhand der Dichtemessung kann dann durch Umrechnung auf den Stärkegehalt geschlossen werden. Zur Umrechnung wurde folgende Formel angewendet:

Die spezifische Dichte wurde durch Wiegen der Knollen sowohl in der Luft als auch in Wasser bestimmt, wobei x die Masse in Luft und y die Masse in Wasser ist. Die spezifische Dichte ergibt sich dann aus x/(x-y). Desweiteren

wurden folgende Beziehungen aus 560 gemessenen Proben gemittelt (Burton W.G. (1989) The Potato. Longman, New York):

% Trockenmasse = 24.182 + [211.04 * (sp. Dichte - 1.0988)]

% Stärke = 17.546 + [199.07 * (sp. Dichte - 1.0988)]

Für die Messungen der Leghemoglobin exprimierenden Pflanzen und der Kontrollpflanzen zu Vergleichszwecken wurden unterschiedlich große Kartoffelknollen verschiedener transgenen Linien sowie der Kontrollpflanzen eingesetzt. Die unterschiedlichen Größen dienen dazu, den beobachteten Effekt in allen Knollengrößen zu reproduzieren.

Die Aufarbeitung bzw. Gewinnung von Stärke aus den Leghemoglobin-exprimierenden Kartoffeln kann nach dem Fachmann bekannten üblichen verfahren z. B. gemäss den Angaben der US Patentanmeldung 2001/0041199 A1, Seite 4, Beispiel 1, erfolgen.

Für die Messung des Stärkegehaltes in den Knollen der Leghemoglobin exprimierenden Pflanzen sowie der Kontrollpflanzen zu Vergleichszwecken wurden die Pflanzen sowohl im Gewächshaus als auch im Foliengewächshaus angezogen. Die verschiedenen Kultivierungen dienen dazu, den beobachteten Effekt unter verschiedenen Klimabedingungen zu reproduzieren. Für die Messungen wurden unterschiedlich große Kartoffelknollen eingesetzt, um den beobachteten Effekt in allen Knollengrößen zu reproduzieren.

In Tabelle 6 sind beispielhaft die Werte einer transgenen Linie im Vergleich zu Kontrollpflanzen gezeigt, die beide im Gewächshaus kultiviert wurden.

Aufgeführt sind jeweils 6 Messungen verschiedener Knollen einer Linie und die resultierenden Mittelwerte und Standardabweichungen. Dabei kann gemäß der Erfindung eine signifikante Erhöhung des Stärkegehalts in der transgenen Linie um 40,77 % nachgewiesen werden. Für weitere Linien werden ähnliche Werte erreicht.

In Abbildung 5 sind die durchschnittlichen Stärkegehalte von vier unabhängigen transgenen Linien im Vergleich zu Kontrollpflanzen gezeigt, die alle Im Zeitraum vom Mai 2003 bis September 2003 im Foliengewächshaus in Golm kultiviert wurden. Die durchschnittlichen Messwerte basieren hierbei auf 334 Knollen (Linie 13), 358 Knollen (Linie 57), 380 Knollen (Linie 45), 384 Knollen (Linie 54) und 151 Knollen (Wildtyp). Gemäß der Erfindung konnte auch unter den im Foliengewächshaus vorherrschenden Klimabedingungen eine signifikante Erhöhung des Stärkegehaltes in den Knollen der Leghemoglobin-exprimierenden transformierten Pflanzen beobachtet werden, so dass die Ergebnisse aus dem Gewächshaus bestätigt werden konnten. Die im Foliengewächshaus während der Anzucht vorherrschenden Temperaturbedingungen sind der Abbildung 6 zu entnehmen.

Tabelle 3. Ölgehalte in T2 Samen transgener Arabidopsis-Linien, die LjLegHb exprimieren, im Vergleich zur Kontrolle. Die Einzelmessungen wurden mit jeweils 5-10 Samen durchgeführt und die entsprechenden Ölgehalte sind als µg Fettsäuren pro Samen angegeben.

Linie	WT	Linie 1.1.9	Linie 3	Linie 5	Linie 4
	6,39	9,86	6,46	8,43	8,64
	9,38	9,79	6,87	9,04	8,99
	9,38	8,98	7,89	8,91	9,18
	7,48	9,72	9,45	8,77	8,43
	9,38	7,78	8,16		6,57
	6,87	7,54	8,30		10,74
	8,44	7,59	6,97		
	6,03	4,31			
	3,93				
Mittelwert	7,48	8,20	7,73	8,79	8,76
Standardabweichung	1,77	1,74	0,96	0,23	1,23
relative Ölzunahme		**9,61**	**3,34**	**17,54**	**17,14**

5

Tabelle 4. Ölgehalte in T2 Samen transgener Arabidopsis-Linien, die AtHb1 exprimieren, im Vergleich zur Kontrolle. Die Einzelmessungen wurden mit jeweils 5-10 Samen durchgeführt und die entsprechenden Ölgehalte sind als µg Fettsäuren pro Samen

Linie	WT	Linie 39	Linie 35	Linie 17	Linie 13
	6,39	12,17	8,13	7,96	8,02
	9,38	11,02	7,48	6,73	9,04
	9,38	10,54	7,82	7,45	8,16
	7,48	10,13	7,89	6,05	8,09
	9,38	9,93	9,72		8,64
	6,87	13,40			
	8,44	11,36			
	6,03	11,22			
	3,93	11,97			
Mittelwert	7,48	11,30	8,21	7,05	8,39
Standardabweichung	1,77	1,03	0,79	0,72	0,39
relative Ölzunahme		**51,16**	**9,78**	**-5,77**	**12,22**

Tabelle 5. Ölgehalte in T2 Samen transgener Arabidopsis-Linien, die AtHb2 exprimieren, im Vergleich zur Kontrolle. Die Einzelmessungen wurden mit jeweils 5-10 Samen durchgeführt und die entsprechenden Ölgehalte sind als µg Fettsäuren pro Samen angegeben.

Linie	WT	Linie 10	Linie 9	Linie 11	Linie 2
	6,39	8,36	9,79	8,26	8,73
	9,38	8,53	8,50	9,66	8,36
	9,38	9,83	9,79	8,23	9,11
	7,48	9,72	9,93	8,98	8,57
	9,38		6,69	11,16	
	6,87		3,06		
	8,44		7,54		
	6,03				
	3,93				
Mittelwert	7,48	9,11	7,90	9,26	8,69
Standardabweichung	1,77	0,67	2,29	1,09	0,27
relative Ölzunahme		**21,86**	**5,64**	**23,80**	**16,25**

Tabelle 6: Bestimmung der spezifischen Dichte einer transgenen Linie im Vergleich zu einer Kontrollpflanze.

	Wildtyp				**LegHb**			
	Luft	Wasser	Wasser/Luft	Dichte	Luft	Wasser	Wasser/Luft	Dichte
Knolle1	5,4500	0,2700	0,0500	1,0521	13,990	1,0000	0,0710	1,0770
Knolle2	8,9900	0,3300	0,0370	1,0381	15,530	1,1000	0,0710	1,0762
Knolle3	25,7900	1,4700	0,0570	1,0604	30,280	2,2000	0,0730	1,0783
Knolle4	30,7000	1,9300	0,0630	1,0671	50,740	3,2300	0,0640	1,0680
Knolle5	37,4500	1,9900	0,0530	1,0561	59,430	4,0600	0,0680	1,0733
Knolle6	80,9300	4,5000	0,0560	1,0589	67,120	4,3600	0,0650	1,0695
Mittelwert				**1,0555**				**1,0737**
STD				0,0090				0,0039
% Stärke	**8,9181**				**12,554**			
Erhöhung in %					**40,77**			

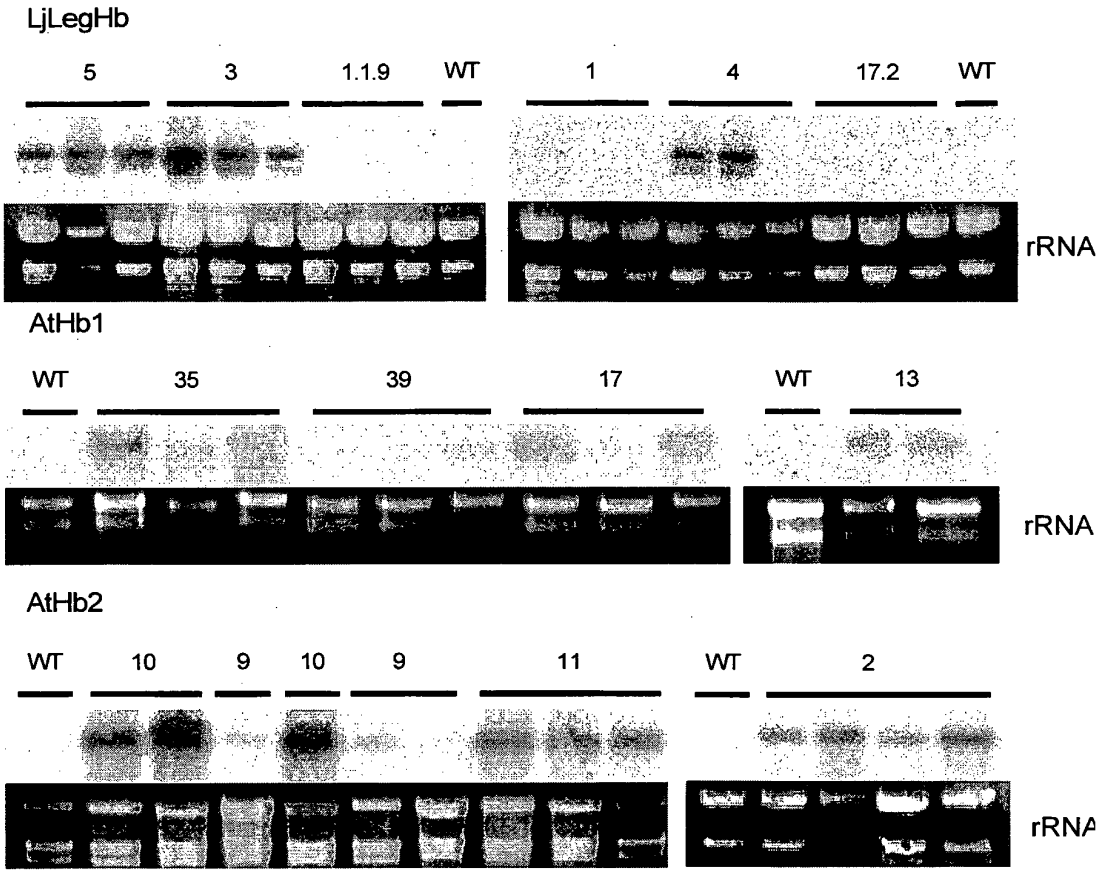


Abbildung 1. Northern Blot-Analysen mit reifenden Samen transgener Arabidopsis-Pflanzen die mit Lotus-Leghemoglobin (LjLegHb), Arabidopsis-Hemoglobin1 (AtHb1) oder Arabidopsis-Hemoglobin2 (AtHb2) transformiert worden waren. Die Menge der eingesetzten Gesamt-RNA erfolgt durch Vergleich der Menge an ribosomaler RNA (rRNA).

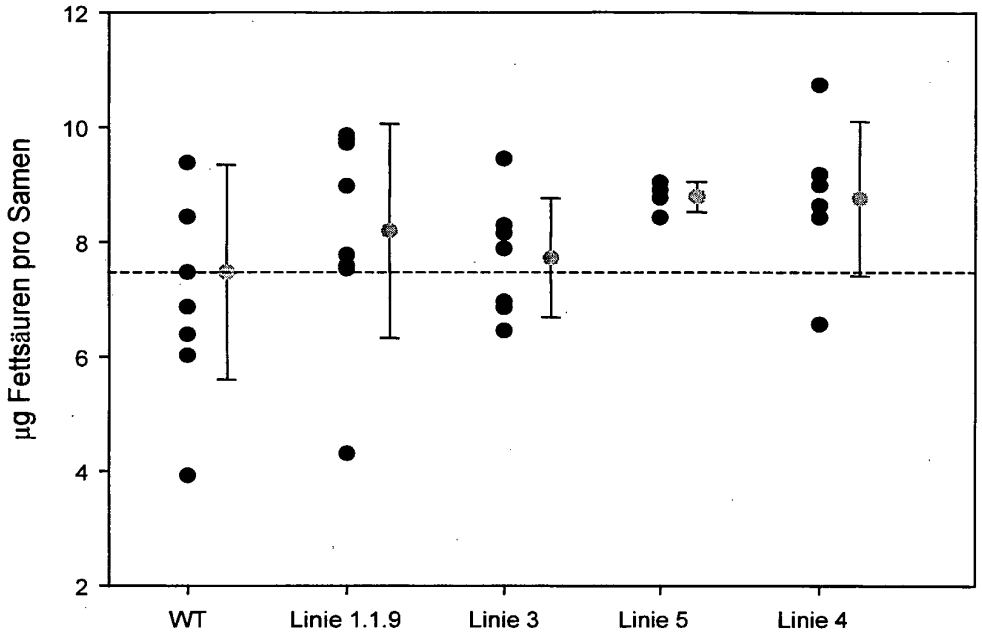


Abbildung 2. Graphische Darstellung der Ölgehalte in T2 Samen transgener Arabidopsis-Linien, die LjLegHb exprimieren, im Vergleich zur Kontrolle. Aufgeführt sind einerseits die Werte aus 4-9 Einzelmessungen mit jeweils 5-10 Samen (●) und andererseits die resultierenden Mittelwerte mit den dazugehörigen Standardabweichungen (⊛).

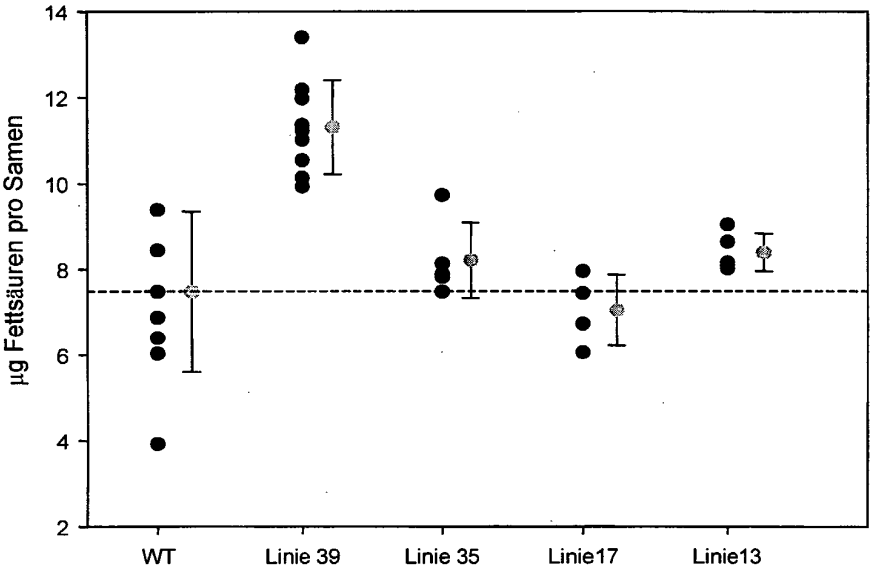


Abbildung 3. Graphische Darstellung der Ölgehalte in T2 Samen transgener Arabidopsis-Linien, die AtHb1 exprimieren, im Vergleich zur Kontrolle. Aufgeführt sind einerseits die Werte aus 4-9 Einzelmessungen mit jeweils 5-10 Samen (•) und andererseits die resultierenden Mittelwerte mit den dazugehörigen Standardabweichungen (⌾).

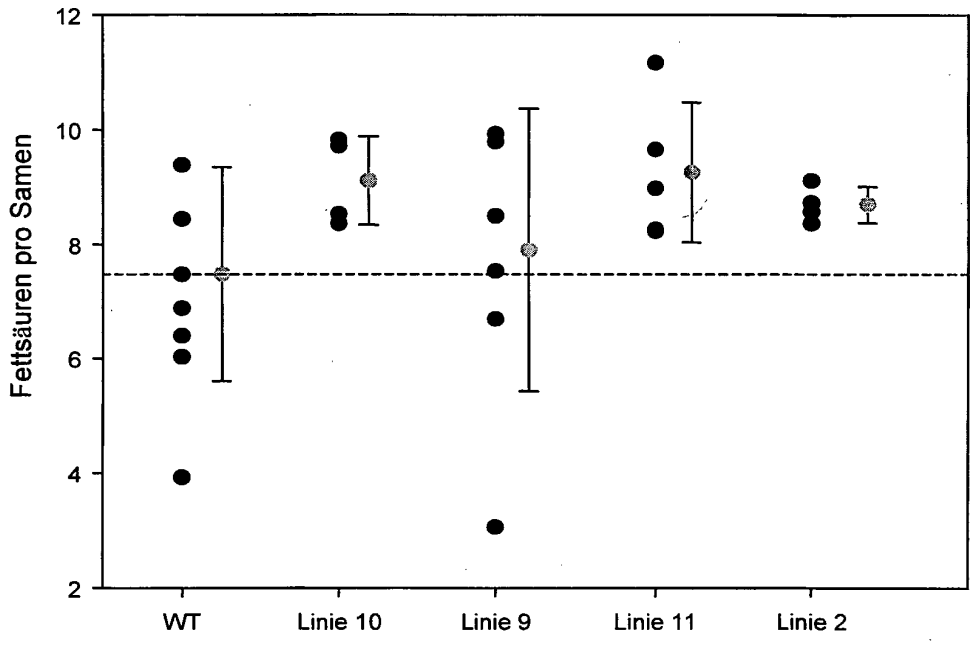


Abbildung 4. Graphische Darstellung der Ölgehalte in T2 Samen transgener Arabidopsis-Linien, die AtHb2 exprimieren, im Vergleich zur Kontrolle. Aufgeführt sind einerseits die Werte aus 4-9 Einzelmessungen mit jeweils 5-10 Samen (•) und andererseits die resultierenden Mittelwerte mit den dazugehörigen Standardabweichungen (⊙).

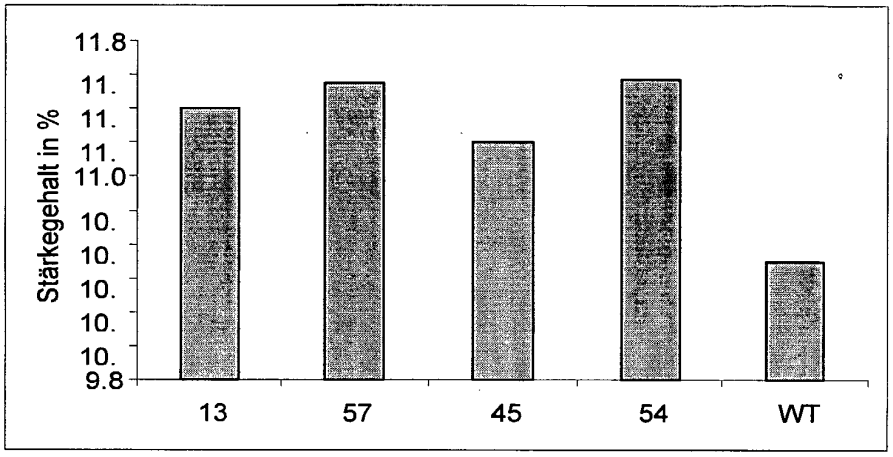


Abbildung 5. Stärkegehalt von Leghemoglobin exprimierenden Kartoffel-Knollen im Vergleich zum Wildtyp. Pflanzen wurden im Foliengewächshaus in Golm im Sommer 2003 gezogen, und abgereifte Knollen geerntet. Der Stärkegehalt wurde durch Bestimmung der Dichte (specific gravity) der Knollen gemessen. Die Daten basieren jeweils auf 334 Knollen (Linie 13), 358 Knollen (Linie 57), 380 Knollen (Linie 45), 384 Knollen (Linie 54) und 151 Knollen (Wildtyp).

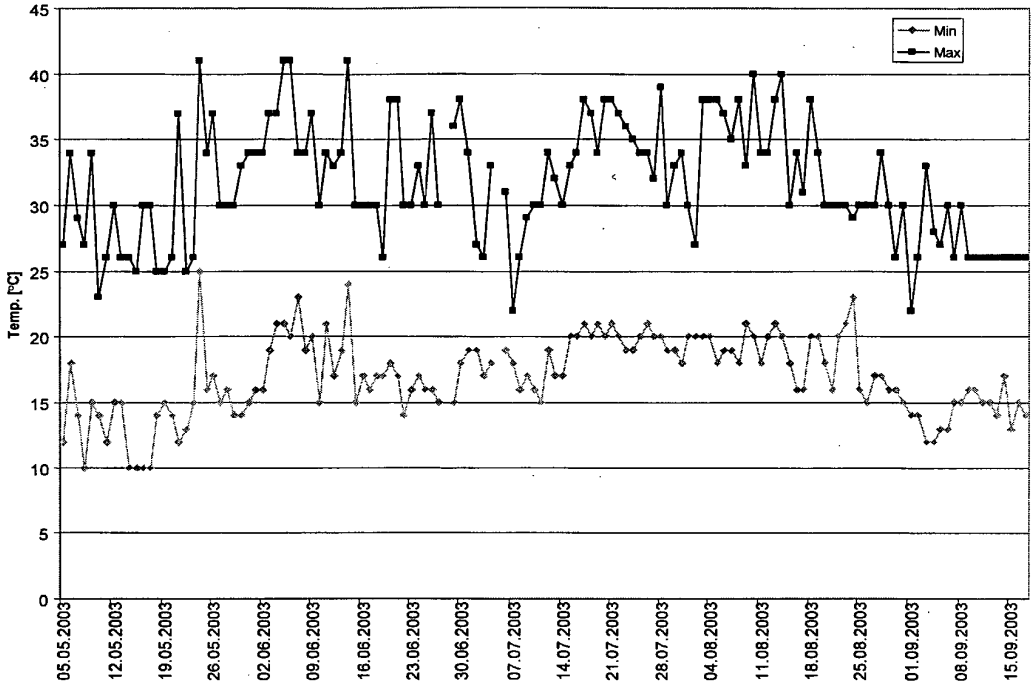


Abbildung 6. Temperaturmessungen im Foliengewächshaus, in dem die transgenen Kartoffelpflanzen, die das Leghemoglobin aus Lotus japonicus überexprimierten, angezogen wurden. Angegeben sind jeweils die Temperaturmaxima und –minima.

5

Patentansprüche

1. Transformierte Pflanze, **dadurch gekennzeichnet, dass** sie mindestens ein Leghemoglobin exprimiert.
2. Transformierte Pflanze nach Anspruch 1, **dadurch gekennzeichnet, dass** sie mindestens eine Sequenz Nr. 1 codierend für ein Leghemoglobin umfasst.
3. Transformierte Pflanze nach Anspruch 2, **dadurch gekennzeichnet, dass** sie eine Sequenz umfasst, die mit der Sequenz Nr. 1 zu ca. 70 % identisch ist.
4. Transformierte Pflanze, **dadurch gekennzeichnet,** dass sie mindestens ein Hemoglobin oder mindestens ein Leghemoglobin und mindestens ein Hemoglobin exprimiert.
5. Transformierte Pflanze nach einem der Ansprüche 1 bis 4, **dadurch gekennzeichnet, dass** das Leghemoglobin und/oder Hemoglobin aus Pflanzen aus der Gruppe bestehend aus *Lupinus luteus*, *Glycine max*, *Medicago sativa*, *Medicago trunculata*, *Phaseolus vulgaris*, *Vicia faba*, *Pisum sativum*, *Vigna unguiculata*, *Lotus japonicus*, *Psophocarpus tetragonolobus*, *Sesbania rostrata*, *Casuarina glauca* und *Canvalaria lineata* ausgewählt wird.
6. Transformierte Pflanze nach einem der Ansprüche 1 bis 5, **dadurch gekennzeichnet, dass** das Leghemoglobin und/oder Hemoglobin aus Lotus japonicus und Arabidopsis thaliana stammt.
7. Transformierte Pflanze nach einem der Ansprüche 1 bis 6, **dadurch gekennzeichnet, dass** sie das Leghemoglobin und/oder Hemoglobin speicherorganspezifisch exprimiert.
8. Transformierte Pflanze nach einem der Ansprüche 1 bis 7, **dadurch gekennzeichnet, dass** sie das mindestens eine Leghemoglobin

und/oder Hemoglobin knollenspezifisch und/oder samenspezifisch exprimiert.

9. Transformierte Pflanze nach einem der Ansprüche 4 bis 8, **dadurch gekennzeichnet, dass** sie mindestens eine der Sequenzen Nr. 3 und 5 codierend für Hemoglobin oder mindestens eine Sequenz Nr. 1 codierend für ein Leghemoglobin und mindestens eine der Sequenzen Nr. 3 und 5 codierend für Hemoglobin umfasst.

10. Transformierte Pflanze nach einem der Ansprüche 4 bis 9, **dadurch gekennzeichnet, dass** sie Sequenzen umfasst, die mit den Sequenzen Nr. 1, 3 und/oder 5 zu ca. 70 % identisch sind.

11. Transformierte Pflanze nach einem der vorhergehenden Ansprüche **dadurch gekennzeichnet, dass** sie Stärke und/oder Öl produziert.

12. Transformierte Pflanze nach einem der vorhergehenden Ansprüche **dadurch gekennzeichnet, dass** es sich um monokotyle Kulturpflanzen, insbesondere der Art Gramineae handelt.

13. Transformierte Pflanze nach einem der Ansprüche 1 bis 12, **dadurch gekennzeichnet, dass** es sich um dikotyledone Kulturpflanzen, insbesondere der Familie Asteraceae, Brassicacea, Compositae, Cruciferae, Cucurbitaceae, Leguminosae, Rubiaceae, Solanaceae, Sterculiaceae, Theaceae oder Umbelliferae handelt.

14. Transformierte Pflanze nach Anspruch 13, **dadurch gekennzeichnet, dass** es sich um Kartoffeln, Arabidopsis thaliana, Soja oder Raps handelt.

15. Nukleotidsequenz gemäß Sequenz Nr. 1 codierend für Leghemoglobin zur Verwendung in einer Pflanze nach einem der Ansprüche 1 bis 14.

16. Genstruktur enthaltend mindestens eine Nukleotidsequenz gemäß Anspruch 15.

17. Vektor enthaltend mindestens eine oder mehrere Nukleotidsequenzen gemäß Anspruch 15 oder eine oder mehrere Genstrukturen gemäß Anspruch 16.

18. Pflanze nach einem der Ansprüche 1 bis 14 enthaltend mindestens eine Genstruktur gemäß Anspruch 16.

19. Pflanze nach einem der Ansprüche 1 bis 14 enthaltend mindestens einen Vektor gemäß Anspruch 17.

20. Nukleotidsequenz gemäß Sequenzen gem. Nr. 3 und 5 codierend für Hemoglobin zur Verwendung in einer Pflanze nach einem der Ansprüche 4 bis 14.

21. Genstruktur enthaltend mindestens eine Nukleotidsequenz gem. Anspruch 20.

22. Vektor enthaltend mindestens eine oder mehrere Nukleotidsequenzen gemäß Anspruch 20 oder eine oder mehrere Genstrukturen gemäß Anspruch 21.

23. Pflanze nach einem der Ansprüche 4 bis 14 enthaltend mindestens eine Genstruktur gemäß Anspruch 21.

24. Pflanze nach einem der Ansprüche 4 bis 14 enthaltend mindestens einen Vektor gemäß Anspruch 24.

25. Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen, **dadurch gekennzeichnet, dass** Pflanzen derart transformiert werden, dass sie mindestens ein Leghemoglobin exprimieren.

26. Verfahren nach Anspruch 25, **dadurch gekennzeichnet, dass** die Pflanzen derart transformiert werden, dass sie mindestens eine Sequenz Nr. 1 codierend für ein Leghemoglobin umfasst.

27. Verfahren nach einem der Ansprüche 25 oder 26, **dadurch gekennzeichnet, dass** die Pflanzen derart transformiert werden, dass

sie eine Sequenz umfassen, die mit der Sequenz Nr. 1 zu ca. 70 % identisch ist.

28. Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen, **dadurch gekennzeichnet, dass** Pflanzen derart transformiert werden, dass sie mindestens Hemoglobin oder ein Leghemoglobin und mindestens ein Hemoglobin exprimieren.

29. Verfahren nach einem der Ansprüche 25 bis 28, **dadurch gekennzeichnet, dass** das Leghemoglobin und Hemoglobin aus Pflanzen aus der Gruppe bestehend aus *Arabidopsis thaliana*, *Lupinus luteus*, *Glycine max*, *Medicago sativa*, *Medicago trunculata*, *Phaseolus vulgaris*, *Vicia faba*, *Pisum sativum*, *Vigna unguiculata*, *Lotus japonicus*, *Psophocarpus tetragonolobus*, *Sesbania rostrata*, *Casuarina glauca* und *Canvalaria lineata* ausgewählt wird.

30. Verfahren nach einem der Ansprüche 25 bis 29, **dadurch gekennzeichnet, dass** das Leghemoglobin und/oder Hemoglobin aus Lotus japonicus Arabidopsis thaliana stammt.

31. Verfahren nach einem der Ansprüche 25 bis 30, **dadurch gekennzeichnet, dass** die Pflanzen derart transformiert werden, dass sie das Leghemoglobin und Hemoglobin speicherorganspezifisch exprimiert.

32. Verfahren nach einem der Ansprüche 25 bis 31, **dadurch gekennzeichnet, dass** die Pflanzen derart transformiert werden, dass sie das Leghemoglobin und Hemoglobin knollenspezifisch und/oder samenspezifisch exprimiert.

33. Verfahren nach einem der Ansprüche 25 bis 32, **dadurch gekennzeichnet, dass** die Pflanzen derart transformiert werden, dass sie mindestens eine Sequenz Nr. 3 und/oder Nr. 5 codierend für Hemoglobin oder mindestens eine Sequenz Nr. 1 codierend für ein

Leghemoglobin und mindestens eine Sequenz Nr. 3 und/oder 5 codierend für Hemoglobin umfasst.

34. Verfahren nach einem der vorhergehenden Ansprüche 26 bis 33, **dadurch gekennzeichnet, dass** die Pflanzen derart transformiert werden, dass sie Sequenzen umfassen, die mit einer der Sequenzen Nr. 1, 3 und/oder 5 zu ca. 70 % identisch sind.

35. Verfahren nach einem der Ansprüche 26 bis 34, **dadurch gekennzeichnet, dass** die Pflanzen derart transformiert werden, dass sie Stärke und/oder Öl produziert.

36. Verfahren nach einem der Ansprüche 26 bis 35, **dadurch gekennzeichnet, dass** monokotyle Kulturpflanzen, insbesondere der Art Gramineae transformiert werden.

37. Verfahren nach einem der vorhergehenden Ansprüche 26 bis 36, **dadurch gekennzeichnet, dass** dikotyledone Kulturpflanzen, insbesondere der Familie Asteraceae, Brassicacea, Compositae, Cruciferae, Cucurbitaceae, Leguminosae, Rubiaceae, Solanaceae, Sterculiaceae, Theaceae oder Umbelliferae transformiert werden.

38. Verfahren nach einem der vorhergehenden Ansprüche 26 bis 37, **dadurch gekennzeichnet, dass** Kartoffeln, Arabidopsis thaliana, Soja oder Raps transformiert werden.

39. Verfahren nach einem der vorhergehenden Ansprüche 26 bis 38, **dadurch gekennzeichnet, dass** mindestens eine Nukleotidsequenz gemäß Anspruch 15 für die Transformation verwendet wird.

40. Verfahren nach einem der Ansprüche 25 bis 39, **dadurch gekennzeichnet, dass** mindestens eine Genstruktur gemäß Anspruch 16 für die Transformation verwendet wird.

41. Verfahren nach einem der Ansprüche 25 bis 40, **dadurch gekennzeichnet, dass** mindestens ein Vektor gemäß Anspruch 17 für die Transformation verwendet wird.

42. Verfahren nach einem der Ansprüche 28 bis 41, **dadurch gekennzeichnet, dass** mindestens eine Nukleotidsequenz gemäß Anspruch 20 für die Transformation verwendet wird.

43. Verfahren nach einem der Ansprüche 28 bis 42, **dadurch gekennzeichnet, dass** mindestens eine Genstruktur gemäß Anspruch 21 für die Transformation verwendet wird.

44. Verfahren nach einem der Ansprüche 28 bis 43, **dadurch gekennzeichnet, dass** mindestens ein Vektor gemäß Anspruch 22 für die Transformation verwendet wird.

45. Verwendung einer Pflanze nach einem der Ansprüche 1 bis 14, 18 oder 19 zur Produktion von Stärke und/oder Öl.

46. Verwendung einer Pflanze nach einem der Ansprüche 1 bis 10, 11 bis 24 zur Herstellung von Stärke und/oder Öl.

SEQUENCE LISTING

<110> Max-Planck-Gesellschaft

<120> Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen

<130> NAE 737/02 PCT

<150> DE 10260707.9
<151> 2002-12-23

<160> 6

<170> PatentIn version 3.2

<210> 1
<211> 444
<212> DNA
<213> Lotus japonicus

<220>
<221> CDS
<222> (1)..(444)

<400> 1
atg ggt ttc act gcg cag caa gag gct cta gtg ggt agc tca tac gaa
48
Met Gly Phe Thr Ala Gln Gln Glu Ala Leu Val Gly Ser Ser Tyr Glu
1 5 10 15

aca ttc aag aaa aac ctt cct acc aac agt gtt ttg ttc tac acc gtt
96
Thr Phe Lys Lys Asn Leu Pro Thr Asn Ser Val Leu Phe Tyr Thr Val

```
                20                  25                  30

ata ttg gag ata gca cca act gca aaa gac atg ttc tcc ttt cta aag
144
Ile Leu Glu Ile Ala Pro Thr Ala Lys Asp Met Phe Ser Phe Leu Lys
        35                  40                  45

gag tct ggg cct aag cat agt cct cag ctc cag gcc cat gct gaa aag
192
Glu Ser Gly Pro Lys His Ser Pro Gln Leu Gln Ala His Ala Glu Lys
    50                  55                  60

gtt ttt gca ctg act cgt gat gct gcc act caa ctc gta gca aaa gga
240
Val Phe Ala Leu Thr Arg Asp Ala Ala Thr Gln Leu Val Ala Lys Gly
65                  70                  75                  80

gaa gtg aca ctt gca gat gcc agc tta ggt gct gtc cac gtt cag aaa
288
Glu Val Thr Leu Ala Asp Ala Ser Leu Gly Ala Val His Val Gln Lys
                85                  90                  95

gcc gtt act gat cct cat ttc gtg gtg gtt aaa gaa gcc ctg ctt caa
336
Ala Val Thr Asp Pro His Phe Val Val Val Lys Glu Ala Leu Leu Gln
            100                 105                 110

aca gta aag gaa gca gtt ggg gcg gac gaa tgg agt gat gac ttg agc
384
Thr Val Lys Glu Ala Val Gly Ala Asp Glu Trp Ser Asp Asp Leu Ser
        115                 120                 125

acc gct tgg gaa gga gca tat gat gga cta gca act gca att aag aag
432
Thr Ala Trp Glu Gly Ala Tyr Asp Gly Leu Ala Thr Ala Ile Lys Lys
```

130 135 140

gca atg ggt taa
444
Ala Met Gly
145

<210> 2
<211> 147
<212> PRT
<213> Lotus japonicus

<400> 2

Met Gly Phe Thr Ala Gln Gln Glu Ala Leu Val Gly Ser Ser Tyr Glu
1 5 10 15

Thr Phe Lys Lys Asn Leu Pro Thr Asn Ser Val Leu Phe Tyr Thr Val
20 25 30

Ile Leu Glu Ile Ala Pro Thr Ala Lys Asp Met Phe Ser Phe Leu Lys
35 40 45

Glu Ser Gly Pro Lys His Ser Pro Gln Leu Gln Ala His Ala Glu Lys
50 55 60

Val Phe Ala Leu Thr Arg Asp Ala Ala Thr Gln Leu Val Ala Lys Gly
65 70 75 80

Glu Val Thr Leu Ala Asp Ala Ser Leu Gly Ala Val His Val Gln Lys
85 90 95

Ala Val Thr Asp Pro His Phe Val Val Val Lys Glu Ala Leu Leu Gln
100 105 110

Thr Val Lys Glu Ala Val Gly Ala Asp Glu Trp Ser Asp Asp Leu Ser
115 120 125

Thr Ala Trp Glu Gly Ala Tyr Asp Gly Leu Ala Thr Ala Ile Lys Lys
130 135 140

Ala Met Gly
145

<210> 3
<211> 483
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (1)..(483)

<400> 3
atg gag agt gaa gga aag att gtg ttc aca gaa gag caa gag gct ctt
48
Met Glu Ser Glu Gly Lys Ile Val Phe Thr Glu Glu Gln Glu Ala Leu
1 5 10 15

```
gta gtg aag tct tgg agt gtc atg aag aaa aac tca gct gaa tta ggt
96
Val Val Lys Ser Trp Ser Val Met Lys Lys Asn Ser Ala Glu Leu Gly
            20                  25                  30

ctc aaa ctc ttc atc aag atc ttt gag att gca cca aca acg aag aag
144
Leu Lys Leu Phe Ile Lys Ile Phe Glu Ile Ala Pro Thr Thr Lys Lys
        35                  40                  45

atg ttc tct ttc ttg aga gac tca cca att cct gct gag caa aat cca
192
Met Phe Ser Phe Leu Arg Asp Ser Pro Ile Pro Ala Glu Gln Asn Pro
    50                  55                  60

aag ctc aag cct cac gca atg tct gtt ttt gtc atg tgt tgt gaa tca
240
Lys Leu Lys Pro His Ala Met Ser Val Phe Val Met Cys Cys Glu Ser
65                  70                  75                  80

gca gta caa ctg agg aaa aca ggg aaa gtt acg gtg agg gag act act
288
Ala Val Gln Leu Arg Lys Thr Gly Lys Val Thr Val Arg Glu Thr Thr
                85                  90                  95

ttg aag aga ctt gga gcc agc cat tct aaa tac ggt gtc gtt gac gaa
336
Leu Lys Arg Leu Gly Ala Ser His Ser Lys Tyr Gly Val Val Asp Glu
            100                 105                 110

cac ttt gag gtg gcc aag tat gca ttg ttg gag acg ata aag gag gca
384
His Phe Glu Val Ala Lys Tyr Ala Leu Leu Glu Thr Ile Lys Glu Ala
        115                 120                 125
```

gtg ccg gag atg tgg tca ccg gag atg aag gtg gct tgg ggt cag gct
432
Val Pro Glu Met Trp Ser Pro Glu Met Lys Val Ala Trp Gly Gln Ala
130 135 140

tat gat cac ctt gtt gct gcc att aaa gct gaa atg aat ctt tcc aac
480
Tyr Asp His Leu Val Ala Ala Ile Lys Ala Glu Met Asn Leu Ser Asn
145 150 155 160

taa
483

<210> 4
<211> 160
<212> PRT
<213> Arabidopsis thaliana

<400> 4

Met Glu Ser Glu Gly Lys Ile Val Phe Thr Glu Glu Gln Glu Ala Leu
1 5 10 15

Val Val Lys Ser Trp Ser Val Met Lys Lys Asn Ser Ala Glu Leu Gly
20 25 30

Leu Lys Leu Phe Ile Lys Ile Phe Glu Ile Ala Pro Thr Thr Lys Lys
35 40 45

Met Phe Ser Phe Leu Arg Asp Ser Pro Ile Pro Ala Glu Gln Asn Pro

50 55 60

Lys Leu Lys Pro His Ala Met Ser Val Phe Val Met Cys Cys Glu Ser
65 70 75 80

Ala Val Gln Leu Arg Lys Thr Gly Lys Val Thr Val Arg Glu Thr Thr
85 90 95

Leu Lys Arg Leu Gly Ala Ser His Ser Lys Tyr Gly Val Val Asp Glu
100 105 110

His Phe Glu Val Ala Lys Tyr Ala Leu Leu Glu Thr Ile Lys Glu Ala
115 120 125

Val Pro Glu Met Trp Ser Pro Glu Met Lys Val Ala Trp Gly Gln Ala
130 135 140

Tyr Asp His Leu Val Ala Ala Ile Lys Ala Glu Met Asn Leu Ser Asn
145 150 155 160

<210> 5
<211> 477
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS

<222> (1)..(477)

<400> 5

atg gga gag att ggg ttt aca gag aag caa gaa gct ttg gtg aag gaa
48
Met Gly Glu Ile Gly Phe Thr Glu Lys Gln Glu Ala Leu Val Lys Glu
1 5 10 15

tcg tgg gag ata ctg aaa caa gac atc ccc aaa tac agc ctt cac ttc
96
Ser Trp Glu Ile Leu Lys Gln Asp Ile Pro Lys Tyr Ser Leu His Phe
20 25 30

ttc tca cag ata ctg gag ata gca cca gca gca aaa ggc ttg ttc tct
144
Phe Ser Gln Ile Leu Glu Ile Ala Pro Ala Ala Lys Gly Leu Phe Ser
35 40 45

ttc cta aga gac tca gat gaa gtc cct cac aac aat cct aaa ctc aaa
192
Phe Leu Arg Asp Ser Asp Glu Val Pro His Asn Asn Pro Lys Leu Lys
50 55 60

gct cat gct gtt aaa gtc ttc aag atg aca tgt gaa aca gct ata cag
240
Ala His Ala Val Lys Val Phe Lys Met Thr Cys Glu Thr Ala Ile Gln
65 70 75 80

ctg agg gag gaa gga aag gtg gta gtg gct gac aca acc ctc caa tat
288
Leu Arg Glu Glu Gly Lys Val Val Val Ala Asp Thr Thr Leu Gln Tyr
85 90 95

tta ggc tca att cat ctc aaa agc ggc gtt att gac cct cac ttc gag
336

Leu Gly Ser Ile His Leu Lys Ser Gly Val Ile Asp Pro His Phe Glu
100 105 110

gtg gtg aaa gaa gct ttg cta agg aca ttg aaa gag ggg ttg ggg gag
384
Val Val Lys Glu Ala Leu Leu Arg Thr Leu Lys Glu Gly Leu Gly Glu
115 120 125

aaa tac aat gaa gaa gtg gaa ggt gct tgg tct caa gct tat gat cac
432
Lys Tyr Asn Glu Glu Val Glu Gly Ala Trp Ser Gln Ala Tyr Asp His
130 135 140

ttg gct tta gcc atc aag acc gag atg aaa caa gaa gag tca taa
477
Leu Ala Leu Ala Ile Lys Thr Glu Met Lys Gln Glu Glu Ser
145 150 155

<210> 6
<211> 158
<212> PRT
<213> Arabidopsis thaliana

<400> 6

Met Gly Glu Ile Gly Phe Thr Glu Lys Gln Glu Ala Leu Val Lys Glu
1 5 10 15

Ser Trp Glu Ile Leu Lys Gln Asp Ile Pro Lys Tyr Ser Leu His Phe
20 25 30

Phe Ser Gln Ile Leu Glu Ile Ala Pro Ala Ala Lys Gly Leu Phe Ser

35 40 45

Phe Leu Arg Asp Ser Asp Glu Val Pro His Asn Asn Pro Lys Leu Lys
50 55 60

Ala His Ala Val Lys Val Phe Lys Met Thr Cys Glu Thr Ala Ile Gln
65 70 75 80

Leu Arg Glu Glu Gly Lys Val Val Val Ala Asp Thr Thr Leu Gln Tyr
85 90 95

Leu Gly Ser Ile His Leu Lys Ser Gly Val Ile Asp Pro His Phe Glu
100 105 110

Val Val Lys Glu Ala Leu Leu Arg Thr Leu Lys Glu Gly Leu Gly Glu
115 120 125

Lys Tyr Asn Glu Glu Val Glu Gly Ala Trp Ser Gln Ala Tyr Asp His
130 135 140

Leu Ala Leu Ala Ile Lys Thr Glu Met Lys Gln Glu Glu Ser
145 150 155

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
15. Juli 2004 (15.07.2004)

PCT

(10) Internationale Veröffentlichungsnummer
WO 2004/057946 A3

(51) Internationale Patentklassifikation[7]: C12N 15/82, C07K 14/415, A01H 5/00

(21) Internationales Aktenzeichen: PCT/EP2003/014774

(22) Internationales Anmeldedatum:
23. Dezember 2003 (23.12.2003)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
102 60 707.9 23. Dezember 2002 (23.12.2002) DE

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: **MAX-PLANCK-GESELLSCHAFT** [DE/DE]; Hofgartenstr. 8, 80539 München (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder *(nur für US)*: **GEIGENBERGER, Peter** [DE/DE]; Quantzstr. 12, 14129 Berlin (DE). **LANGER, Anke** [DE/DE]; Sellostr. 27, 14471 Potsdam (DE). **VIGEOLAS, Helene** [FR/DE]; Am alten Mörtelwerk 14, 14469 Potsdam (DE). **STITT NIGEL, Marc** [DE/DE]; Grosse Weinmeisterstr. 22a, 14469 Potsdam (DE). **THOMAS VAN DONGEN, Joost** [NL/DE]; Leipziger Str. 52, 14473 Potsdam (DE). **UDVARDI, Michael** [DE/DE]; Golmer Fichten 32, 14476 Golm (DE).

(74) Anwalt: **FITZNER, Uwe**; Lintorfer Strasse 10, 40878 Ratingen (DE).

(81) Bestimmungsstaaten *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Bestimmungsstaaten *(regional)*: ARIPO Patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:
— *mit internationalem Recherchenbericht*
— *vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen*

(88) Veröffentlichungsdatum des internationalen Recherchenberichts: 2. September 2004

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(54) **Title:** METHOD FOR ALTERING THE CONTENT OF RESERVE SUBSTANCES IN PLANTS

(54) **Bezeichnung:** VERFAHREN ZUR VERÄNDERUNG DES GEHALTS VON SPEICHERSTOFFEN IN PFLANZEN

(57) **Abstract:** The invention relates to a method for altering the content of reserve substances in plants during which leghemoglobin and/or hemoglobin expressing transformed plants are used. The invention also relates to corresponding plants and to the use thereof.

(57) **Zusammenfassung:** Die Erfindung betrifft ein Verfahren zur Veränderung des Gehalts von Speicherstoffen in Pflanzen, wobei Leghemoglobin und/oder Hemoglobin exprimierende transformierte Pflanzen eingesetzt werden, entsprechende Pflanzen sowie deren Verwendung.

WO 2004/057946 A3

INTERNATIONAL SEARCH REPORT

Intern Application No
PCT/EP 03/14774

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/82 C07K14/415 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS, EMBASE, CAB Data, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99/02687 A (JARDINAUD MARIE FRANCOISE ;PEACOCK WILLIAM JAMES (AU); COMMW SCIEN) 21 January 1999 (1999-01-21) see SEQ ID NO: 1 and 2 (pp. 50-54) page 33; claims 1-8,15,18-21,24-28,36; figures 3,5,8; example 1	1,3-14, 20-24, 28-38, 42-44
X	US 6 372 961 B1 (TARCZYNSKI MITCHELL C ET AL) 16 April 2002 (2002-04-16) column 1 - column 2; claims 1-4,8,9,11,13,14,17,18,22-31,33; examples 1,2 column 11 - column 12	1,3-8, 10-14, 25, 27-29, 31,32, 34-38

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 June 2004	25/06/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Oderwald, H

2

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

Intern[illegible] l Application No
PCT/EP 03/14774

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BARATA R M ET AL: "Targeting of the soybean leghemoglobin to tobacco chloroplasts: Effects on aerobic metabolism in transgenic plants" PLANT SCIENCE (SHANNON), vol. 155, no. 2, 29 June 2000 (2000-06-29), pages 193-202, XP002283115 ISSN: 0168-9452 abstract page 194, paragraph 2 page 196, paragraph 4 page 197, paragraph 4 - page 198, paragraph 1	1,3-5, 13,25, 28,29,37
X	STOUGAARD, J. ET AL: "Expression of a complete soybean leghemoglobin gene in root nodules of transgenic Lotus corniculatus." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, PUBLISHER: WASHINGTON, D.C.: THE ACADEMY. CODEN: PNASA6;ISSN: 0027-8424, vol. 84, no. 16, 1987, pages 5754-5757, XP002283117 abstract page 5754, paragraph 2	1,3-5,13
X	WO 98/12913 A (BAILEY JAMES E ;BULOW LEIF (SE)) 2 April 1998 (1998-04-02) cited in the application abstract; claims 1-3,6,9,14,32; figure 4; example 1 page 1 - page 4 page 6 - page 8 page 12 - page 15 page 23	1,3-5,7, 8,11-14, 25, 27-29, 31,32, 35-38
X	DATABASE EMBL 'Online! EBI, Hinxton; 18 November 2000 (2000-11-18), UCHIUMI T ET AL: "Genomic leghemoglobin genes from Lotus japonicus" XP002283119 Database accession no. AB042718 abstract	15-19

-/--

2

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/EP 03/14774

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! EBI, Hinxton; 6 November 1997 (1997-11-06), TREVASKIS B ET AL: "Two hemoglobin genes in Arabidopsis thaliana: The evolutionary origins of leghemoglobins" XP002283120 Database accession no. U94999 abstract	20-22
X	-& TREVASKIS B ET AL: PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA PUBLISHER: WASHINGTON, D.C.: NATIONAL ACADEMY OF SCIENCES, CODEN: PNASA6;ISSN: 0027-8424, vol. 94, October 1997 (1997-10), pages 12230-12234, XP002283114 abstract page 12233, last paragraph	23,24

2

Form PCT/ISA/210 (continuation of second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

on on patent family members

Intern: ıl Application No
PCT/EP 03/14774

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9902687	A	21-01-1999	AU	8199598 A	08-02-1999
			WO	9902687 A1	21-01-1999
			CA	2295202 A1	21-01-1999
			EP	1007681 A1	14-06-2000
			JP	2001509385 T	24-07-2001
US 6372961	B1	16-04-2002	NONE		
WO 9812913	A	02-04-1998	US	5959187 A	28-09-1999
			AU	730663 B2	08-03-2001
			AU	4502297 A	17-04-1998
			BR	9711556 A	18-01-2000
			CA	2266626 A1	02-04-1998
			EP	0955804 A1	17-11-1999
			HU	9904237 A2	28-04-2000
			JP	2001501098 T	30-01-2001
			WO	9812913 A1	02-04-1998

Form PCT/ISA/210 (patent family annex) (January 2004)

INTERNATIONALER RECHERCHENBERICHT

Intern es Aktenzeichen
PCT/ EP 03/14774

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 7 C12N15/82 C07K14/415 A01H5/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 7 C12N C07K A01H

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)
EPO-Internal, WPI Data, PAJ, BIOSIS, EMBASE, CAB Data, Sequence Search

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	WO 99/02687 A (JARDINAUD MARIE FRANCOISE ;PEACOCK WILLIAM JAMES (AU); COMMW SCIEN) 21. Januar 1999 (1999-01-21) see SEQ ID NO: 1 and 2 (pp. 50-54) Seite 33; Ansprüche 1-8,15,18-21,24-28,36; Abbildungen 3,5,8; Beispiel 1	1,3-14, 20-24, 28-38, 42-44
X	US 6 372 961 B1 (TARCZYNSKI MITCHELL C ET AL) 16. April 2002 (2002-04-16) Spalte 1 - Spalte 2; Ansprüche 1-4,8,9,11,13,14,17,18,22-31,33; Beispiele 1,2 Spalte 11 - Spalte 12	1,3-8, 10-14, 25, 27-29, 31,32, 34-38

-/--

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

[X] Siehe Anhang Patentfamilie

° Besondere Kategorien von angegebenen Veröffentlichungen :
"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
7. Juni 2004	25/06/2004

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Oderwald, H

2

Formblatt PCT/ISA/210 (Blatt 2) (Januar 2004)

INTERNATIONALER RECHERCHENBERICHT

Interna les Aktenzeichen
PCT/EP 03/14774

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	BARATA R M ET AL: "Targeting of the soybean leghemoglobin to tobacco chloroplasts: Effects on aerobic metabolism in transgenic plants" PLANT SCIENCE (SHANNON), Bd. 155, Nr. 2, 29. Juni 2000 (2000-06-29), Seiten 193-202, XP002283115 ISSN: 0168-9452 Zusammenfassung Seite 194, Absatz 2 Seite 196, Absatz 4 Seite 197, Absatz 4 - Seite 198, Absatz 1	1,3-5, 13,25, 28,29,37
X	STOUGAARD, J. ET AL: "Expression of a complete soybean leghemoglobin gene in root nodules of transgenic Lotus corniculatus." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, PUBLISHER: WASHINGTON, D.C.: THE ACADEMY. CODEN: PNASA6;ISSN: 0027-8424, Bd. 84, Nr. 16, 1987, Seiten 5754-5757, XP002283117 Zusammenfassung Seite 5754, Absatz 2	1,3-5,13
X	WO 98/12913 A (BAILEY JAMES E ;BULOW LEIF (SE)) 2. April 1998 (1998-04-02) in der Anmeldung erwähnt Zusammenfassung; Ansprüche 1-3,6,9,14,32; Abbildung 4; Beispiel 1 Seite 1 - Seite 4 Seite 6 - Seite 8 Seite 12 - Seite 15 Seite 23	1,3-5,7, 8,11-14, 25, 27-29, 31,32, 35-38
X	DATABASE EMBL 'Online! EBI, Hinxton; 18. November 2000 (2000-11-18), UCHIUMI T ET AL: "Genomic leghemoglobin genes from Lotus japonicus" XP002283119 Database accession no. AB042718 Zusammenfassung	15-19

-/--

2

INTERNATIONALER RECHERCHENBERICHT

Intern es Aktenzeichen
PCT/ER 03/14774

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	DATABASE EMBL 'Online! EBI, Hinxton; 6. November 1997 (1997-11-06), TREVASKIS B ET AL: "Two hemoglobin genes in Arabidopsis thaliana: The evolutionary origins of leghemoglobins" XP002283120 Database accession no. U94999 Zusammenfassung	20-22
X	-& TREVASKIS B ET AL: PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA PUBLISHER: WASHINGTON, D.C.: NATIONAL ACADEMY OF SCIENCES, CODEN: PNASA6;ISSN: 0027-8424, Bd. 94, Oktober 1997 (1997-10), Seiten 12230-12234, XP002283114 Zusammenfassung Seite 12233, letzter Absatz _____	23,24

2

Formblatt PCT/ISA/210 (Fortsetzung von Blatt 2) (Januar 2004)

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichung[illegible] zur selben Patentfamilie gehören

Interna[illegible]s Aktenzeichen
PCT/[illegible] 03/14774

Im Recherchenbericht angeführtes Patentdokument		Datum der Veröffentlichung	Mitglied(er) der Patentfamilie		Datum der Veröffentlichung
WO 9902687	A	21-01-1999	AU	8199598 A	08-02-1999
			WO	9902687 A1	21-01-1999
			CA	2295202 A1	21-01-1999
			EP	1007681 A1	14-06-2000
			JP	2001509385 T	24-07-2001
US 6372961	B1	16-04-2002	KEINE		
WO 9812913	A	02-04-1998	US	5959187 A	28-09-1999
			AU	730663 B2	08-03-2001
			AU	4502297 A	17-04-1998
			BR	9711556 A	18-01-2000
			CA	2266626 A1	02-04-1998
			EP	0955804 A1	17-11-1999
			HU	9904237 A2	28-04-2000
			JP	2001501098 T	30-01-2001
			WO	9812913 A1	02-04-1998

Formblatt PCT/ISA/210 (Anhang Patentfamilie) (Januar 2004)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 November 2004 (18.11.2004)

PCT

(10) International Publication Number
WO 2004/099405 A1

(51) International Patent Classification[7]: C12N 15/00, 15/63, A01H 5/00

(21) International Application Number: PCT/US2003/009209

(22) International Filing Date: 11 April 2003 (11.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **VENTRIA BIOSCIENCE** [US/US]; 4110 North Freeway Boulevard, Sacramento, CA 95834 (US).

(71) Applicants and
(72) Inventors: **HUANG, Ning** [CN/US]; 1226 Caricia Drive, Davis, CA 95616 (US). **RODRIQUEZ, Raymond, L.** [US/US]; 3017 Anza Avenue, Davis, CA 95616 (US). **HAGIE, Frank, E.** [US/US]; 4400 Truxel Road, #191, Sacramento, CA 95834 (US). **STALKER, David, M.** [US/US]; 870 W. Southwood Drive, Woodland, CA 95695 (US).

(74) Agents: **BERMAN, Richard, J.** et al.; Arent Fox Kintner Plotkin & Kahn, PLLC, 1050 Connecticut Avenue, NW, Suite 400, Washington, DC 20036-5339 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/099405 A1

(54) Title: HUMAN BLOOD PROTEINS EXPRESSED IN MONOCOT SEEDS

(57) Abstract: The invention is directed to blood proteins produced in monocot seeds and isolated therefrom for use in therapeutic compositions, and to methods of making these isolated blood proteins and to therapeutic compositions comprising them.

HUMAN BLOOD PROTEINS EXPRESSED IN MONOCOT SEEDS

Field of the Invention

The present invention relates to human blood proteins produced in the seeds of monocot plants for use in making human and animal topical compositions and human therapeutic compositions.

Background Of The Invention

Many human blood proteins are in short or limited supply due to the larger quantities required of the protein for positive therapeutic effect or possibly also due to the larger demand of these proteins by the world population of patients having the particular condition. It is also advantageous to produce blood proteins, normally extracted from blood products, from an alternative source such as crop plants. Production of blood proteins from plants mitigates contamination of the blood protein fraction with human viruses and other disease causative agents found in human or animal blood product fractions.

Blood proteins such as hemoglobin, alpha-1-antitrypsin ("AAT"), fibrinogen, human serum albumin, thrombin, antibodies, blood coagulation factors (e.g. Factors V-XIII), and others are known to have therapeutic potential for a number of human conditions.

Hemoglobin is the major blood component molecule transporting oxygen to cells. Mammmalian hemoglobins are tetrameric proteins made up of two α-like polypeptide subunits and two non-α (usually β, γ, or δ) subunits. These subunits differ in primary amino acid sequence, but have similar secondary and tertiary structures. Each globin subunit has associated with it, by noncovalent interaction, a Fe^{2+}-porphyrin complex known as a heme group, to which oxygen binds. The predominant hemoglobin in adult erythrocytes is $\alpha 2\beta 2$, known as hemoglobin A_1 (HbA). Each hemoglobin tetramer has a molecular weight of 64kD and each α-like and β-like chain has a molecular weight of approximately 15.7kD (141 amino acids) and 16.5kD (146 amino acids) respectively.

AAT belongs to the class of serpin inhibitors and is one of the major protease inhibitors in human plasma. AAT is a single 394 amino acid polypeptide having an approximate molecular mass of 52kD, and contains about 15% carbohydrate in the native human form of the molecule. Concentrations of AAT in human plasma range from 1000-3000 mg/L and in human milk range from 100 to 400 mg/L. Its primary

physiological role is the inhibition neutrophil elastase, with an insufficiency leading to the development of pulmonary emphysema. Excess production of elastase activity leads to emphysema, hepatitis and a variety of skin disorders. While the binding affinity of AAT is highest for human neutrophil elastase, it also has affinity for pancreatic proteases such as chymotrypsin and trypsin. The current primary source for the treatment of AAT deficiency is isolating AAT from human blood plasma.

Fibrinogen is involved in the blood coagulation cascade and is converted to fibrin by its interaction with the natural clotting agent thrombin. Fibrin is the major component of blood clotting. Mature human fibrinogen consists of two pairs of three independent polypeptide chains (α, β and γ) that are linked together by 29 intra- and intermolecular disulfide bonds forming a native protein of 340kD and is present in human plasma at an approximate concentration of 2500 mg/L. Three-dimentional structural analysis of independent fibrinogen domains has provided detailed structural features giving important clues to human fibrinogen's multifunctional role. The fibrinogen polypeptides are approximately 72kD (α), 52kD (β) and 48kD (γ) respectively with the β polypeptide chain determining native molecule assembly. The structure of fibrinogen features a number of structural and functional domains containing multiple binding sites that facilitate interactions with itself, other proteins, certain cell types and allow fibrinogen to participate ina number of important physiological processes including blood coagulation, inflammation, angiogenesis, wound closure, artheriogenesis and tumorigenesis. Fibrin formation from a clotting standpoint is mediated by the interaction of native fibrinogen with its natural clotting agents Factor XIII and thrombin in the presence of blood soluble calcium.

Albumin is a transport protein molecule that carries out many functions in mammalian serum biology, notably that of a carrier of hormones and other soluble ligands from site to site, and other activities that contribute largely to general mammalian biochemistry. Human serum albumin ("HSA") is also the major protein component of blood being actively present at serum concentrations of approximately 30,000-50,000 mg/L. HSA is a single polypeptide chain of 66.5kD that is initially synthesized as a prepro-albumin molecule in the liver and released from the endoplasmic reticulum after N-terminal and C-terminal Golgi processing. The resultant mature protein is 585 amino acids in length. It has been shown that the natural

preprosequnce of HSA can function in correct protein targeting/processing across a plant plasma membrane in transgenic tobacco leaves (Sijmons *et al*, 1990).

Prothrombin, a plasma glycoprotein, is the zymogen of the serine protease thrombin that catalyzes the conversion of fibrinogen to fibrin as well as several other reactions that may be important for blood coagulation. Prothrombin is a single polypeptide chain approximately 72,000 molecular weight in size. The complete human thrombin cDNA consists of 622 amino acid residues and includes a leader sequence of 36 amino acid residues. Active thrombin has an apparent molecular weight of 36,000 and is made up of two disulfide-linked polypeptide chains resulting from prothrombin cleavage. The proteolytic events leading to *in vitro* activation and conversion of human prothrombin to active thrombin have been extremely well characterized.

Factors V-XIII are proteins (mostly proteases in their active states) that are involved in the 'intrinsic pathway' of the classical casade mechanism for blood coagulation. The majority of these molecules exist as precursors that are processed in an ordered sequence of transformations from inactive to catalytically active forms. Factor V is proaccelerin (the accelerator globulin) while Factor VI is the activated form of Factor V. Factor VII is proconvertin, the plasma thromboplastin component, while Factors VIII (antihemophilic factor) and IX (Christmas antihemophilic factor) are both associated with the hemophilia disease state. Factors X (Stuart-Power factor), XI (plasma thromboplastin anticedent) and XII (Hageman factor) are all involved with the maturation/stabilization of thrombin. Factor XIII (fibrin stabilizing factor) is a plasma transglutaminase directly acting on fibrin during the clotting process. All these Factors are present at relatively low in serum plasma (0.001 to 50 mg/L). Other protein factors also involved in the blood coagulation cascade include Fletcher Factor (prekallikrein), Fitzgerald factor (kininogen) and von Willebrand Factor.

Immunoglobulins (antibodies) present in humans act to confer resistance to a variety of pathogens to which a patient may have been exposed. Immunoglobulin molecules account for 15-20% of the mass in human serum and consist predominantly of IgG, IgM and IgA-type antibodies involved in fighting various infections that invade the blood system and potentially the rest of the body. IgG type antibodies are the most prevelant and exist at a serum concentration of between 6-18 g/L. The blood system also serves as a carrier directing these molecules to specific areas of the body to combat resulting infections and potential oncogenic targets. Mature antibodies consist

of two polypeptides (light and heavy chains) that must be expressed in eqimolar amounts and come together to form functional entities. The light chain (~25kD) is a protein of ~210-240 amino acids in length while the heavy chain (~50kD) is a protein of ~450-460 amino acids in length. Both light and heavy chains carry signal peptides for processing and secretion into the blood stream. Expression of monoclonal antibodies in plants is of particular interest, because it requires the expression of two genes, synthesis of two proteins and coerrect assembly of the tetrameric protein to result in a functional antibody.

Initial studies of antibodies in plants focused on the IgG antibodiy class (Hiatt *et al,* 1989; Hiatt and Ma, 1992), but later studies explored the *in planta* expression of complex antibody molecules such as secretory IgA antibodies (4 genes) and more complex antibody forms (Ma *et al*, 1995; Vine *et al,* 2001).

U.S. Patent Nos. 6,417,429, 5,959,177, 5,639,947 and 5,202,422, all related patents, disclose the production of antibody molecules in transgenic tobacco plant leaves.

U.S. Patent No. 6,303,341 discloses the production of immunoglobulins containing protection proteins in tobacco plant leaves, stems, flowers and roots.

Published U.S. Patent Application U.S. 2002/0174453 discloses the production of antibodies in the plastids of tobacco plants.

Published U.S. Patent Application U.S. 2002/0046418 discloses a controlled environment agriculture bioreactor for the commercial production of heterologous proteins in transgenic plants. The specification discloses that production of mammalian blood proteins can be achieved. Example 7 discloses the production of human blood factors in the leaves of potato, tobacco and alfalfa plants.

U.S. Patent No. 6,344,600 discloses the production of hemoglobin and myoglobin in tobacco plant leaves. Example X discloses the extraction and partial purification of recombinant hemoglobin from tobacco seeds. The expression was obtained by transformation of the coexpression plasmid pBIOC59, which was constructed to allow targeting in the chloroplasts, and contained for this purpose the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase of *Pisum sativum* L. Expression in seeds was reported to be at a maximum level of 0.05% recombinant hemoglobin relative to the total soluble proteins extracted.

Example XI of the '600 patent discloses the construction of plasmids containing one of the α or β chains of hemoglobin allowing constitutive expression or expression in the albumin in maize seeds. According to this disclosure, the constitutive or albumin-specific expression of the hemoglobin chains required the following regulatory sequences: one of three promoters allowing a constitutive expression ((i) the rice actin promoter followed by the rice actin intron, contained in the plasmid pAct1-F4; (ii) the 35S double constitutive promoter of cauliflower mosaic virus; or (iii) the promoter of the maize γ-zein gene contained in the plasmid pγ63) and one of two terminators ((i) the 35S polyA terminator; or (ii) the NOS polyA terminator). No experiment or data is provided regarding transformation or expression of these plasmids in maize or maize seeds.

U.S. Patent No. 5,767,363 discloses the use of a seed-specific promoter derived from ACP of *Brassica napus*, to affect and vary the expression of seed oils in rape and tobacco plants. The specification generically discloses that the seed-specific promoter can be used for the expression of pharmaceutical proteins, such as blood factors or human serum albumin, however no experimental data whatsoever is presented in this regard.

Daniell *et al.* (2001) is a review article discussing recent developments in the field of medical molecular farming, including the production of antibodies and proteins in plants.

None of these patents or publications discloses the production of human blood proteins in monocot seeds in high yield. It is desirable to provide for the production of human blood proteins in high yield free from contaminating source agents in order to provide the patient population with sufficient supply of these proteins for use in treating humans with conditions treatable by administration of a particular blood protein.

Summary Of The Invention

In one aspect, the invention includes a method of producing a recombinant human blood protein in monocot plant seeds, comprising the steps of:

(a) transforming a monocot plant cell with a chimeric gene comprising

(i) a promoter from the gene of a maturation-specific monocot plant storage protein,

(ii) a first DNA sequence, operably linked to said promoter, encoding a monocot plant seed-specific signal sequence capable of targeting a polypeptide linked thereto to a monocot plant seed endosperm cell, and

(iii) a second DNA sequence, linked in translation frame with the first DNA sequence, encoding a human blood protein, wherein the first DNA sequence and the second DNA sequence together encode a fusion protein comprising an N-terminal signal sequence and the human blood protein;

(b) growing a monocot plant from the transformed monocot plant cell for a time sufficient to produce seeds containing the human blood protein; and

(c) harvesting the seeds from the plant,

wherein the human blood protein constitutes at least 3.0% of the total soluble protein in the harvested seeds.

The invention also includes a purified human blood protein obtained by the method. Preferably, the human blood protein comprises one or more plant glycosyl groups.

The invention also provides a monocot plant seed product, preferably selected from whole seed, flour, extract and malt, prepared from the harvested seeds obtained by the method of the invention. Preferably, the human blood protein constitutes at least 3.0% of the total soluble protein in the seed product.

The invention further provides a composition comprising a purified human blood protein, preferably comprising at least one plant glycosyl group, and at least one pharmaceutically acceptable excipient or nutrient, wherein the human blood protein is produced in a monocot plant containing a nucleic acid sequence encoding the human blood protein and is purified from seed harvested from the monocot plant. The nutrient is from a source other than the monocot plant. The formulation can be used for parenteral, enteric, inhalation, intranasal or topical delivery.

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read in conjunction with the accompanying drawings and claims.

Brief Description of the Figures

Figure 1 shows plasmids with constructs containing three codon-optimized genes encoding the fibrinogen polypeptides α (pAPI 398), β (pAPI 417) and γ (pAPI 327) (SEQ

ID NO: 1-3), each under the control of the rice glutelin promoter Gt1. These plasmids, including a plasmid (not shown) containing the hygromycin selectable marker, were bombarded into embryogenic rice callus to create transgenic rice plants expressing these three genes in mature rice seeds.

Figure 2 shows a Western blot analysis of transgenic rice lines expressing individual subunits of human fibrinogen. Lane 1, positive control, purified native human fibrinogen (obtained from the Red Cross) showing all three polypeptide chains; Lane 2, extract from Tapei 309, a non transgenic rice variety; Lane 3, molecular weight standard; Lane 4, rice seed extract expressing fibrinogen α chain; Lane 5, rice seed extract expressing fibrinogen β chain; Lane 6, rice seed extract expressing fibrinogen γ chain. Total protein extract of rice seeds was performed in 2% SDS, 1M urea, 1% βMe and PBS pH 7.4. Fibrinogen polypeptides were detected using antibody recognizing all three chains or individual chains only.

Figure 3 shows the simultraneous expression of the three fibrinogen polypeptide chains (α, β and γ) in transgenic rice seeds and analyzed via Western blot analysis. Fibrinogen polypeptides and protein aggregates were detected using antibody recognizing all three chains. Figure 3A indicates total protein extracted from rice seeds under non-denaturing conditions (350 mM NaCl, PBS pH 7.4, 0.01% Tween-20/Trition X-100/CHAPS) and run on a non-denaturing 10% acrylamide gel. Lane 1, 1 μg purified human fibrinogen; Lanes 2 & 3, extracts from Tapei 309, a non-transgenic rice variety; Lane 4, molecular weight markers; Lanes 5 & 7, extracts from two transgenic rice lines where 1.0% βMe was included in the extraction buffer; lanes 6 & 8, extracts from two transgenic rice lines without βMe in the extraction buffer. Lanes 6 & 8 show large protein aggregates that were extracted under non-denaturing conditions from the transgenic lines that run at the approximate position of complexed native human fibrinogen. Figure 3B indicates total protein extracted from rice seeds in 2% SDS, 1M urea, 1% βMe and PBS pH 7.4, and run on SDS-PAGE. Lane 1, positive control, native human fibrinogen (obtained from the Red Cross) showing all three polypeptide chains; Lane 2, molecular weight standards; Lanes 3-5, three independent transgenic rice lines expressing all three fibrinigen polypeptides.

Figure 4 shows the plasmid pAPI 250 expressing the codon-optimized gene for alpha-1-antitrypsin (AAT) (SEQ ID NO: 5) under the control of the rice glutelin promoter

Gt1. This plasmid, along with a plasmid (not shown) containing the hygromycin selectable marker gene, was bombarded into embryogenic rice callus to create transgenic rice plants expressing AAT in mature rice seeds.

Figure 5 shows Coomassie brilliant blue staining of aqueous phase extraction of transgenic rice grains expressing human recombinant AAT. Both untransformed (rice var. Kitaake) and transgenic rice seeds (~10 pooled R1 seed from five individual transgenic plants) were ground with PBS pH 7.4 buffer. The resulting extract was spun at 14,000 rpm at 4° C for 10 min. Supernatant was collected and ~20 μg of this soluble protein extract was resuspended in sample loading buffer, and loaded onto a precast SDS-PAGE gel. Lane 1, molecular weight protein markers; Lane 2, purified non-recombinant human AAT; Lane 3, extract from control non-transformed Kitaake variety. Between lanes 2 and 3, the results from the extracts of the five individual transgenic plants are shown.

Figure 6 shows Western blot analysis of recombinant human AAT expressed in transgenic rice grains. The R1 pooled seed soluble protein extracts (~10 μg total protein) from seven independent transgenic rice transformants were prepared as described in Figure 5 above, separated by SDS-PAGE gel and then blotted onto a nitrocellulose filter. The identification of AAT expressed in rice seeds was carried out by Western analysis using anti-AAT antibody. Lane 1, molecular weight protein markers; Lanes 2 & 3, 1 μg and 2 μg, respectively, of purified non-recombinant human AAT; Lanes 4 & 5, control, non-transgenic rice extract (var. Kitaake). The final seven lanes show the results from the extracts of the seven individual transgenic plants. Extracts from two of the seven transgenic lines did not express AAT. The shift in gel mobility between the non-recombinant human and recombinant rice-expressed forms is due to the type and glycosylation differences in the human and recombinant rice-expressed proteins.

Figure 7 shows activity of purified recombinant AAT (rAAT) obtained from rice extracts against purified porcine pancreatic elastase (PPE) as determined by Coomassie staining and Western blot analysis. The activity of rAAT is demonstrated by a band shift assay involving the AAT protease substrate, elastase. AAT samples from human and rice ectracts were incubated with equal number of moles of PPE at 37°C for 15 min. Negative control for band shift assay was prepared with the AAT samples

incubated with equal volume of PPE added. Lane MW refers to molecular weight markers. Figure 7A: Lane 1, purified non-recombinant AAT from human plasma; Lane 2, purified AAT from human plasma + PPE; Lane 3, soluble protein extract containing AAT from transgenic rice seed; Lane 4, protein extract containing AAT from transgenic rice seed + PPE; Lane 5, non-transformed rice seed extract; Lane 6, non-transformed rice seed extract + PPE. Figure 7B shows a shifted band in Lanes 1, 2 and 3. The shifted band, a complex between PPE and an AAT fragment is confirmed to contain AAT by Western blot analysis. The lanes in Figure 7B are analogous to those in Figure 7A.

Figure 8 depicts AAT derived from rice cell extracts purified initially through Con-A and DEAE Sepharose respectively, then loaded onto an octyl Sepharose column. Octyl Sepharose is the final purification step and separates active AAT from an inactivated form of the protein. Lane 1, molecular weight markers; Lane 2, 2 μg purified non-recombinant human AAT as a standard; Lane 3, pooled eluate from the DEAE Sepharose column. The remaining columns show the flow-through and the eluate from the octyl Sepharose column. Approximately 50 μL from each column fraction was loaded onto an SDS-PAGE gel and the proteins visualized by Coomassie staining. Octyl Sepharose flow-through shows the inactive AAT protein while the eluate resolves active AAT.

Figure 9A depicts an AAT association rate constant for activity of purified recombinant AAT against PPE determined (as described by the procedure in Figure 7) using non-recombinant human AAT as a control. Data were generated by Coomassie protein staining and Western blot analysis, as described in Figure 7. Figure 9B depicts the thermostability of plant-derived recombinant AAT versus native human AAT determend by the PPE inhibition assay.

Figure 10 shows the plasmid pAPI 9 for expression of codon-optimized human serum albumin (HSA) (SEQ ID NO: 4) under the control of the rice Amy1A promoter/signal peptide. This plasmid is useful for the expression of HSA in germinated rice seeds.

Figure 11 shows the expression of HSA in transgenic rice seeds. Pooled seed from transgenic rice line 3-11-2 were imbibed in water for 24 hours, then 2 μM gibberelic acid (GA) was added. Seed samples were extracted at 24, 48, 72, and 120 hours post

GA addition and soluble proteins were extracted and prepared for Western analysis. 15 μg of soluble protein were loaded onto each lane along with protein isolated from the non-transfromed negative control line TP309. The blot was probed with monoclonal antibody prepared against HSA.

Detailed Description of the Invention

Unless otherwise indicated, all terms used herein have the meanings given below or are generally consistent with same meaning that the terms have to those skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., Ausubel FM *et al.* (1993) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., and Gelvin and Schilperoot , eds. (1997) Plant Molecular Biology Manual, Kluwer Academic Publishers, The Netherlands for definitions and terms of the art.

The polynucleotides of the invention may be in the form of RNA or in the form of DNA, and include messenger RNA, synthetic RNA and DNA, cDNA, and genomic DNA. The DNA may be double-stranded or single-stranded, and if single-stranded may be the coding strand or the non-coding (anti-sense, complementary) strand.

The term "stably transformed" with reference to a plant cell means the plant cell has a non-native (heterologous) nucleic acid sequence integrated into its genome which is maintained through two or more generations.

By "host cell" is meant a cell containing a vector and supporting the replication and/or transcription and/or expression of the heterologous nucleic acid sequence. Preferably, according to the invention, the host cell is a monocot plant cell. Other host cells may be used as secondary hosts, including bacterial, yeast, insect, amphibian or mammalian cells, to move DNA to a desired plant host cell.

A "plant cell" refers to any cell derived from a plant, including undifferentiated tissue (*e.g.*, callus) as well as plant seeds, pollen, propagules, embryos, suspension cultures, meristematic regions, leaves, roots, shoots, gametophytes, sporophytes and microspores.

The term "mature plant" refers to a fully differentiated plant.

The term "seed product" includes, but is not limited to, seed fractions such as de-hulled whole seed, flour (seed that has been de-hulled by milling and ground into a powder) a seed extract, preferably a protein extract (where the protein fraction of the flour has been separated from the carbohydrate fraction), malt (including malt extract or malt syrup) and/or a purified protein fraction derived from the transgenic grain.

The term "biological activity" refers to any biological activity typically attributed to that protein by those of skill in the art.

The term "blood protein" refers to one or more proteins, or biologically active fragments thereof, found in normal human blood, including, without limitation, hemoglobin, alpha-1-antitrypsin, fibrinogen, human serum albumin, prothrombin/thrombin, antibodies, blood coagulation factors (Factor V, Factor VI, Factor VII, Factor VIII, Factor IX, Factor X, Factor XI, Factor XII, Factor XIII, Fletcher Factor, Fitzgerald Factor and von Willebrand Factor), and biologically active fragments thereof.

The term "non-nutritional" refers to a pharmaceutically acceptable excipient which does not as its primary effect provide nutrition to the recipient. Preferably, it may provide one of the following services to an enterically delivered formulation, including acting as a carrier for a therapeutic protein, protecting the protein from acids in the digestive tract, providing a time-release of the active ingredients being delivered, or otherwise providing a useful quality to the fomulation in order to administer to the patient the blood proteins.

"Monocot seed components" refers to carbohydrate, protein, and lipid components extractable from monocot seeds, typically mature monocot seeds.

"Seed maturation" refers to the period starting with fertilization in which metabolizable reserves, *e.g.*, sugars, oligosaccharides, starch, phenolics, amino acids, and proteins, are deposited, with and without vacuole targeting, to various tissues in the seed (grain), *e.g.*, endosperm, testa, aleurone layer, and scutellar epithelium, leading to grain enlargement, grain filling, and ending with grain desiccation.

"Maturation-specific protein promoter" refers to a promoter exhibiting substantially upregulated activity (greater than 25%) during seed maturation.

"Heterologous DNA" refers to DNA which has been introduced into plant cells from another source, or which is from a plant source, including the same plant source, but which is under the control of a promoter that does not normally regulate expression of the heterologous DNA.

"Heterologous protein" is a protein encoded by a heterologous DNA.

A "signal sequence" is an N- or C-terminal polypeptide sequence which is effective to localize the peptide or protein to which it is attached to a selected intracellular or extracellular region. Preferably, according to the invention, the signal sequence targets the attached peptide or protein to a location such as an endosperm cell, more preferably an endosperm-cell organelle, such as an intracellular vacuole or other protein storage body, chloroplast, mitochondria, or endoplasmic reticulum, or extracellular space, following secretion from the host cell.

Expression vectors for use in the present invention are chimeric nucleic acid constructs (or expression vectors or cassettes), designed for operation in plants, with associated upstream and downstream sequences.

In general, expression vectors for use in practicing the invention include the following operably linked components that constitute a chimeric gene: a promoter from the gene of a maturation-specific monocot plant storage protein, a first DNA sequence, operably linked to the promoter, encoding a monocot plant seed-specific signal sequence (such as an N-terminal leader sequence or a C-terminal trailer sequence) capable of targeting a polypeptide linked thereto to an endosperm cell, preferably an endosperm-cell organelle, such as a protein storage body, and a second DNA sequence, linked in translation frame with the first DNA sequence, encoding a human blood protein. The signal sequence is preferably cleaved from the human blood protein in the plant cell.

The chimeric gene, in turn, is typically placed in a suitable plant-transformation vector having (i) companion sequences upstream and/or downstream of the chimeric gene which are of plasmid or viral origin and provide necessary characteristics to the vector to permit the vector to move DNA from bacteria to the desired plant host; (ii) a selectable marker sequence; and (iii) a transcriptional termination region generally at the opposite end of the vector from the transcription initiation regulatory region.

Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of plant host cells. The promoter region is chosen to be regulated in a manner allowing for induction under seed-maturation conditions. In one aspect of this embodiment of the invention, the expression construct includes a promoter which exhibits specifically upregulated activity during seed maturation. Promoters for use in the invention are typically derived from cereals such

as rice, barley, wheat, oat, rye, corn, millet, triticale or sorghum. Examples of such promoters include the maturation-specific promoter region associated with one of the following maturation-specific monocot plant storage proteins: rice glutelins, oryzins, and prolamines, barley hordeins, wheat gliadins and glutelins, maize zeins and glutelins, oat glutelins, and sorghum kafirins, millet pennisetins, and rye secalins. Exemplary regulatory regions from these genes are exemplified by SEQ ID NOS: 6-14. Other promoters suitable for expression in maturing seeds include the barley endosperm-specific B1-hordein promoter, GluB-2 promoter, Bx7 promoter, Gt3 promoter, GluB-1 promoter and Rp-6 promoter, particularly if these promoters are used in conjunction with transcription factors.

Of particular interest is the expression of the nucleic acid encoding a human blood protein from a promoter that is preferentially expressed in plant seed tissue. Examples of such promoter sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Exemplary preferred promoters include a glutelin (Gt1) promoter, as exemplified by SEQ ID NO: 6, which effects gene expression in the outer layer of the endosperm, and a globulin (Glb) promoter, as exemplfieid by SEQ ID NO: 7, which effects gene expression in the center of the endosperm. Promoter sequences for regulating transcription of gene coding sequences operably linked thereto include naturally-occurring promoters, or regions thereof capable of directing seed-specific transcription, and hybrid promoters, which combine elements of more than one promoter. Methods for construction such hybrid promoters are well known in the art.

In some cases, the promoter is native to the same plant species as the plant cells into which the chimeric nucleic acid construct is to be introduced. In other embodiments, the promoter is heterologous to the plant host cell.

Alternatively, a seed-specific promoter from one type of monocot may be used regulate transcription of a nucleic acid coding sequence from a different monocot or a non-cereal monocot.

In addition to encoding the protein of interest, the expression cassette or heterologous nucleic acid construct includes DNA encoding a signal peptide that allows processing and translocation of the protein, as appropriate. Exemplary signal sequences are those sequences associated with the monocot maturation-specific genes: glutelins, prolamines, hordeins, gliadins, glutenins, zeins, albumin, globulin, ADP

glucose pyrophosphorylase, starch synthase, branching enzyme, Em, and lea. Exemplary sequences encoding a signal peptide for a protein storage body are identified herein as SEQ ID NOS: 15-21.

In one preferred embodiment, the method is directed toward the localization of proteins in an endosperm cell, preferably an endosperm-cell organelle, such as a protein storage body, mitochondrion, endoplasmic reticulum, vacuole, chloroplast or other plastidic compartment. For example, when proteins are targeted to plastids, such as chloroplasts, in order for expression to take place the construct also employs the use of sequences to direct the gene product to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, when the gene of interest is not directly inserted into the plastid, the expression construct additionally contains a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, (Smeekens *et al.,* 1986; Wasmann *et al.,* 1986; Von Heijne *et al.,* 1991, U.S. patents 4,940,835 and 5,728,925; . Additional transit peptides for the translocation of the protein to the endoplasmic reticulum (ER) (Chrispeels, 1991; Vitale and Chrispeels, 1992), nuclear localization signals (Shieh *et al.,* 1993; Varagona *et al.,* 1992)or vacuole (Raikhel and Chrispeels 1992; Bednarek and Raikel, 1992; also see U.S. Patent No 5,360,726) may also find use in the constructs of the present invention.

Another exemplary class of signal sequences are sequences effective to promote secretion of heterologous protein from aleurone cells during seed germination, including the signal sequences associated with alpha-amylase, protease, carboxypeptidase, endoprotease, ribonuclease, DNase/RNase, (1-3)-beta-glucanase, (1-3)(1-4)-beta-glucanase, esterase, acid phosphatase, pentosamine, endoxylanase, β-xylopyranosidase, arabinofuranosidase, beta-glucosidase, (1-6)-beta-glucanase, perioxidase, and lysophospholipase.

Since many protein storage proteins are under the control of a maturation-specific promoter, and this promoter is operably linked to a signal sequence for targeting to a protein body, the promoter and signal sequence can be isolated from a single protein-storage gene, then operably linked to a blood protein in the chimeric gene construction. One preferred and exemplary promoter-signal sequence is from the rice

Gt1 gene, having an exemplary sequence identified by SEQ ID NO: 6. Alternatively, the promoter and leader sequence may be derived from different genes. One preferred and exemplary promoter-signal sequence combination is the rice Glb promoter linked to the rice Gt1 leader sequence, as exemplified by SEQ ID NO: 7.

Preferably, expression vectors or heterologous nucleic acid constructs designed for operation in plants comprise companion sequences upstream and downstream to the expression cassette. The companion sequences are of plasmid or viral origin and provide necessary characteristics to the vector to permit the vector to move DNA from a secondary host to the plant host, such as, sequences containing an origin of replication and a selectable marker. Typical secondary hosts include bacteria and yeast.

In one embodiment, the secondary host is *E. coli*, the origin of replication is a ColE1-type, and the selectable marker is a gene encoding ampicillin resistance. Such sequences are well known in the art and are commercially available as well (*e.g.*, Clontech, Palo Alto, Calif.; Stratagene, La Jolla, CA).

The transcription termination region may be taken from a gene where it is normally associated with the transcriptional initiation region or may be taken from a different gene. Exemplary transcriptional termination regions include the NOS terminator from *Agrobacterium* Ti plasmid and the rice α-amylase terminator.

Polyadenylation tails may also be added to the expression cassette to optimize high levels of transcription and proper transcription termination, respectively. Polyadenylation sequences include, but are not limited to, the *Agrobacterium* octopine synthetase signal, or the nopaline synthase of the same species.

Suitable selectable markers for selection in plant cells include, but are not limited to, antibiotic resistance genes, such as, kanamycin (*npt*II), G418, bleomycin, hygromycin, chloramphenicol, ampicillin, tetracycline, and the like. Additional selectable markers include a bar gene which codes for bialaphos resistance; a mutant EPSP synthase gene which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil; a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance; and a methotrexate resistant DHFR gene.

The particular marker gene employed is one which allows for selection of transformed cells as compared to cells lacking the DNA which has been introduced. Preferably, the selectable marker gene is one which facilitates selection at the tissue culture stage, *e.g.*, a kanamyacin, hygromycin or ampicillin resistance gene.

The vectors of the present invention may also be modified to include intermediate plant transformation plasmids that contain a region of homology to an *Agrobacterium tumefaciens* vector, a T-DNA border region from *Agrobacterium tumefaciens*, and chimeric genes or expression cassettes (described above). Further, the vectors of the invention may comprise a disarmed plant tumor inducing plasmid of *Agrobacterium tumefaciens*.

In general, a selected nucleic acid sequence is inserted into an appropriate restriction endonuclease site or sites in the vector. Standard methods for cutting, ligating and transformation into a secondary host cell, known to those of skill in the art, are used in constructing vectors for use in the present invention. (See generally, Maniatis *et al.*, Ausubel *et al.*, and Gelvin *et al.*, *supra*.)

Plant cells or tissues are transformed with expression constructs (heterologous nucleic acid constructs, *e.g.*, plasmid DNA into which the gene of interest has been inserted) using a variety of standard techniques. Effective introduction of vectors in order to facilitate enhanced plant gene expression is an important aspect of the invention. It is preferred that the vector sequences be stably transformed, preferably integrated into the host genome.

The method used for transformation of host plant cells is not critical to the present invention. The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available. Any technique that is suitable for the target host plant may be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to, as a strand of DNA, in a plasmid, or in an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to calcium-phosphate-DNA co-precipitation, electroporation, microinjection, *Agrobacterium*-mediated transformation, liposome-mediated transformation, protoplast fusion or microprojectile bombardment (Christou, 1992; Sanford *et al.,* 1993). The skilled artisan can refer to the literature for details and select suitable techniques for use in the methods of the present invention.

When *Agrobacterium* is used for plant cell transformation, a vector is introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA

for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) is inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, examples of which are described in the literature, for example pRK2 or derivatives thereof. See, for example, Ditta *et al.*, 1980 and EPA 0 120 515. Alternatively, one may insert the sequences to be expressed in plant cells into a vector containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride *and Summerfelt* 1990, wherein the pRiHRI (Jouanin, *et al.,* 1985), origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA is one or more selectable marker coding sequences which allow for selection of transformed Agrobacterium and transformed plant cells. A number of antibiotic resistance markers have been developed for use with plant cells, these include genes inactivating antibiotics such as kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, with a particular marker preferred depending on the particular host and the manner of construction.

For *Agrobacterium*-mediated transformation of plant cells, explants are incubated with *Agrobacterium* for a time sufficient to result in infection, the bacteria killed, and the plant cells cultured in an appropriate selection medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of the recombinant protein produced by the plants.

There are a number of possible ways to obtain plant cells containing more than one expression construct. In one approach, plant cells are co-transformed with a first and second construct by inclusion of both expression constructs in a single transformation vector or by using separate vectors, one of which expresses desired

genes. The second construct can be introduced into a plant that has already been transformed with the first expression construct, or alternatively, transformed plants, one having the first construct and one having the second construct, can be crossed to bring the constructs together in the same plant.

In a preferred embodiment, the plants used in the methods of the present invention are derived from members of the taxonomic family known as the *Gramineae*. This includes all members of the grass family of which the edible varieties are known as cereals. The cereals include a wide variety of species such as wheat (*Triticum sps.*), rice (*Oryza sps.*) barley (*Hordeum sps.*) oats, (*Avena sps.*) rye (*Secale sps.*), corn (maize) (*Zea sps.*) and millet (*Pennisettum sps.*). In practicing the present invention, preferred grains are rice, wheat, maize, barley, rye and triticale, and most preferred is rice.

In order to produce transgenic plants that express human blood protein in seeds, monocot plant cells or tissues derived from them are transformed with an expression vector comprising the coding sequence for a human blood protein. The transgenic plant cells are cultured in medium containing the appropriate selection agent to identify and select for plant cells which express the heterologous nucleic acid sequence. After plant cells that express the heterologous nucleic acid sequence are selected, whole plants are regenerated from the selected transgenic plant cells. Techniques for regenerating whole plants from transformed plant cells are generally known in the art. Transgenic plant lines, *e.g.,* rice, wheat, corn or barely, can be developed and genetic crosses carried out using conventional plant breeding techniques.

Transformed plant cells are screened for the ability to be cultured in selective media having a threshold concentration of a selective agent. Plant cells that grow on or in the selective media are typically transferred to a fresh supply of the same media and cultured again. The explants are then cultured under regeneration conditions to produce regenerated plant shoots. After shoots form, the shoots are transferred to a selective rooting medium to provide a complete plantlet. The plantlet may then be grown to provide seed, cuttings, or the like for propagating the transformed plants. The method provides for efficient transformation of plant cells and regeneration of transgenic plants, which can produce a recombinant human blood protein.

The expression of the recombinant human blood protein may be confirmed using standard analytical techniques such as Western blot, ELISA, PCR, HPLC, NMR, or

mass spectroscopy, together with assays for a biological activity specific to the particular protein being expressed.

A purified blood protein recombinantly produced in a plant cell, preferably substantially free of contaminants of the host plant cell, and preferably comprising at least one plant glycosyl group is also provided by the invention. The plant glycosyl groups, while identifying that the blood protein was produced in a plant, does not significantly impair the biological activity of the blood protein in any of the applied therapeutic contexts (preferably less than 25% loss of activity, more preferably less than 10% loss of activity, as compared to a corresponding non-recombinant human blood protein). Typically, in accordance with some embodiments of the invention, the human blood protein constitutes at least about 0.5%, at least about 1.0% or at least about 2.0% of the total soluble protein in the seeds harvested from the transgenic plant. In a preferred embodiment, however, protein expression is much higher than previously reported, i.e., at least about 3.0%, which makes commercial production quite feasible. Advantageously, protein expression is at least about 5.0%, at least about 10%, at least about 15%, at least about 20%, at least about 30%, or even at least about 40% of total soluble protein.

The invention includes plant seed product prepared from the harvested seeds. Preferably, the human blood protein constitutes at least about 3.0% of the total soluble protein in the seed product, more preferably at least about 5.0%, and most preferably at least about 10.0%. As shown in the figures, the expression of human blood proteins in rice grains, represented by AAT, the three fibrinogen polypeptides and HSA represent at least about 10% of total soluble protein.

The present invention also provides compositions comprising human blood proteins produced recombinantly in the seeds of monocot plants, and methods of making such compositions. In practicing the invention, a human blood protein is produced in the seeds or grain of transgenic plants that express the nucleic acid coding sequence for the human blood protein. After expression, the blood protein may be provided to a patient in substantially unpurified form (i.e., at least 20% of the composition comprises plant material), or the blood protein may be isolated or purified from a product of the mature seed (e.g., flour, extract, malt or whole seed, etc.) and formulated for delivery to a patient.

Such compositions can comprise a formulation for the type of delivery intended. Delivery types can include, e.g. parenteral, enteric, inhalation, intranasal or topical delivery. Parenteral delivery can include, e.g. intravenous, intramuscular, or suppository. Enteric delivery can include, e.g. oral administration of a pill, capsule, or other formulation made with a non-nutritional pharmaceutically-acceptable excipient, or a composition with a nutrient from the transgenic plant, for example, in the grain extract in which the protein is made, or from a source other than the transgenic plant. Such nutrients include, for example, salts, saccharides, vitamins, minerals, amino acids, peptides, and proteins other than the human blood protein. Intranasal and inhalant delivery systems can include spray or aerosol in the nostrils or mouth. Topical delivery can include, e.g. creams, topical sprays, or salves. Preferably, the composition is substantially free of contaminants of the transgenic plant, preferably containing less than 20% plant material, more preferably less than 10%, and most preferably, less than 5%. The preferable route of administration is enteric, and preferably the composition is non-nutrititional.

The blood protein can be purified from the seed product by a mode including grinding, filtration, heat, pressure, salt extraction, evaporation, or chromatography.

The human blood proteins produced in accordance with the invention also include all variants thereof, whether allelic variants or synthetic variants. A "variant" human blood protein-encoding nucleic acid sequence may encode a variant human blood protein amino acid sequence that is altered by one or more amino acids from the native blood protein sequence, preferably at least one amino acid substitution, deletion or insertion. The nucleic acid substitution, insertion or deletion leading to the variant may occur at any residue within the sequence, as long as the encoded amino acid sequence maintains substantially the same biological activity of the native human blood protein. In another embodiment, the variant human blood protein nucleic acid sequence may encode the same polypeptide as the native sequence but, due to the degeneracy of the genetic code, the variant has a nucleic acid sequence altered by one or more bases from the native polynucleotide sequence.

The variant nucleic acid sequence may encode a variant amino acid sequence that contains a "conservative" substitution, wherein the substituted amino acid has structural or chemical properties similar to the amino acid which it replaces and physicochemical amino acid side chain properties and high substitution frequencies in

homologous proteins found in nature (as determined, *e.g.*, by a standard Dayhoff frequency exchange matrix or BLOSUM matrix). In addition, or alternatively, the variant nucleic acid sequence may encode a variant amino acid sequence containing a "non-conservative" substitution, wherein the substituted amino acid has dissimilar structural or chemical properties to the amino acid it replaces. Standard substitution classes include six classes of amino acids based on common side chain properties and highest frequency of substitution in homologous proteins in nature, as is generally known to those of skill in the art and may be employed to develop variant human blood protein-encoding nucleic acid sequences.

As will be understood by those of skill in the art, in some cases it may be advantageous to use a human blood protein-encoding nucleotide sequences possessing non-naturally occurring codons. Codons preferred by a particular eukaryotic host can be selected, for example, to increase the rate of human blood protein expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence. As an example, it has been shown that codons for genes expressed in rice are rich in guanine (G) or cytosine (C) in the third codon position (Huang *et al.*, 1990). Changing low G + C content to a high G + C content has been found to increase the expression levels of foreign protein genes in barley grains (Horvath *et al.,* 2000). The blood protein encoding genes employed in the present invention were synthesized by Operon Technologies (Alameda, CA) based on the rice gene codon bias (Huang *et al.*, 1990) along with the appropriate restriction sites for gene cloning. These 'codon-optimized' genes were linked to regulatory/secretion sequences for seed-directed monocot expression and these chimeric genes then inserted into the appropriate plant transformation vectors.

A human blood protein-encoding nucleotide sequence may be engineered in order to alter the human blood protein coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the human blood protein by a cell.

Heterologous nucleic acid constructs may include the coding sequence for a given human blood protein (i) in isolation; (ii) in combination with additional coding sequences; such as fusion protein or signal peptide, in which the human blood protein coding sequence is the dominant coding sequence; (iii) in combination with non-coding

sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; and/or (iv) in a vector or host environment in which the human blood protein coding sequence is a heterologous gene.

Depending upon the intended use, an expression construct may contain the nucleic acid sequence encoding the entire human blood protein, or a portion thereof. For example, where human blood protein sequences are used in constructs for use as a probe, it may be advantageous to prepare constructs containing only a particular portion of the human blood protein encoding sequence, for example a sequence which is discovered to encode a highly conserved human blood protein region.

The invention provides, in one embodiment, a seed composition containing a flour, extract, or malt obtained from mature monocot seeds and one or more seed-produced human blood proteins in unpurified form. Isolating the blood proteins from the flour can entail forming an extract composition by milling seeds to form a flour, extracting the flour with an aqueous buffered solution, and optionally, further treating the extract to partially concentrate the extract and/or remove unwanted components. In a preferred method, mature monocot seeds, such as rice seeds, are milled to a flour, and the flour then suspended in saline or in a buffer, such as Phosphate Buffered Saline ("PBS"), ammonium bicarbonate buffer, ammonium acetate buffer or Tris buffer. A volatile buffer or salt, such as ammonium bicarbonate or ammonium acetate may obviate the need for a salt-removing step, and thus simplify the extract processing method.

The flour suspension is incubated with shaking for a period typically between 30 minutes and 4 hours, at a temperature between 20-55°C. The resulting homogenate is clarified either by filtration or centrifugation. The clarified filtrate or supernatant may be further processed, for example by ultrafiltration or dialysis or both to remove contaminants such as lipids, sugars and salt. Finally, the material maybe dried, *e.g.*, by lyophilization, to form a dry cake or powder. The extract combines advantages of high blood-protein yields, essentially limiting losses associated with protein purification.

In general, the protein once produced in a product of a mature seed can be further purified by standard methods known in the art, such as by filtration, affinity column, gel electrophoresis, and other such standard procedures. The purified protein can then be formulated as desired for delivery to a human patient. More than one

protein can be combined for the therapeutic formulation. The protein may be purified and used in biomedical applications requiring a non-food administration of the protein.

The following examples illustrate but are not intended in any way to limit the invention.

EXAMPLE 1

Production of transgenic rice encoding AAT and fibrinogen polypeptides

The basic procedures of particle bombardment-mediated rice transformation and plant regeneration were carried out as described by Huang *et al.,* 2001. Rice variety TP309 seeds were dehusked, sterilized in 50% (v/v) commercial bleach for 25 min and washed with sterile water. The sterilized seeds were placed on rice callus induction medium (RCI) plates containing [N6 salts (Sigma), B5 vitamins (Sigma), 2mg/l 2,4-D and 3% sucrose]. The rice seeds were incubated for 10 days to induce callus formation. Primary callus was dissected from the seeds and placed on RCI for 3 weeks. This was done twice more to generate secondary and tertiary callus which was used for bombardment and continued subculture. A callus of 1-4mm diameter was placed in a 4cm circle on RCI with 0.3M mannitol 0.3M sorbitol for 5-24 hrs prior to bombardment. Microprojectile bombardment was carried out using the Biolistic PDC-1000/He system (Bio-Rad). The procedure requires 1.5 mg gold particles (60 ug/ml) coated with 2.5 ug DNA. DNA-coated gold particles were bombarded into rice calli with a He pressure of 1100psi. After bombardment, the callus was allowed to recover for 48 hrs and then transferred to RCI with 30 mg/l hygromycin B for selection and incubated in the dark for 45 days at 26°C. Transformed calli were selected and transferred to RCI (minus 2,4-D) containing 5 mg/l ABA, 2 mg/l BAP, 1 mg/l NAA and 30mg/l hygromycin B for 9-12 days. Transformed calli were transferred to regeneration medium consisting of RCI (minus 2,4-D), 3 mg/l BAP, and 0.5 mg/l NAA without hygromycin B and cultured under continuous lighting conditions fro 2-4 weeks. Regenerated plantlets (1-3 cm high) were transferred to rooting medium whose concentration was half that of MS medium (Sigma) plus 1% sucrose and 0.05 mg/l NAA. After 2 weeks on rooting medium, the plantlets developed roots and the shoots grew to about 10 cm. The plants were transferred to a 6.5 x 6.5 cm pots containing a mix of 50% commercial soil (Sunshine #1) and 50% soil from rice fields. The plants were covered by a plastic container to maintain nearly 100% humidity and grown under continuous light for 1 week. The transparent plastic cover

was slowly shifted over a 1 day period to gradually reduce humidity and water and fertilizers added as necessary. When the transgenic R0 plants were approximately 20 cm in height, they were transferred to a greenhouse where they grew to maturity.

Individual R1 seed grains from the individual R0 regenerated plants were dissected into embryos and endosperms. Expression levels of recombinant blood proteins (AAT and fibrinogen poypeptides) in the isolated rice endosperms were determined. Embryos from the individual R1 grains with high recombinant protein expression were sterilized in 50% bleach for 25 min and washed with sterile distilled water. Sterilized embryos were placed in a tissue culture tube containing ½ MS basal salts with the addition of 1% sucrose and 0.05 mg/l NAA. Embryos were germinated and plantlets having ~7 cm shoots and healthy root systems were obtained in about 2 weeks. Mature R1 plants were obtained as regenerants.

EXAMPLE 2

Production of rice extract containing recombinant blood proteins and its use in parenteral and enteric formulations

General procedure for production of rice extract

Transgenic rice containing heterologous polypeptides can be converted to rice extracts by either a dry milling or wet milling process. In the dry milling process, transgenic paddy rice seeds containing the heterologous polypeptides were dehusked with a dehusker. The rice was grounded into a fine flour though a dry milling process, for example, in one experiment, at speed 3 of a model 91 Kitchen Mill from K-TEC. Phosphate buffered saline ("PBS"), containing 0.135 N NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, at pH 7.4, with or without additional NaCl, such as 0.35 N NaCl, was added to the rice flour. In some experiments, approximately 10 ml of extraction buffer was used for each 1 g of flour. In other experiments, the initial flour/buffer ratio varied over a range such as 1 g/40 ml to 1 g/10 ml. The mixture was incubated at room temperature with gentle shaking for 1 hr. In other experiments, the incubation temperature was lower or higher, such as from about 22°C to about 60°C, and the incubation time was longer or shorter, such as from about 10 minutes to about 24 hr. A Thermolyne VariMix platform mixer set at high speed was used to keep the particulates suspended.

In place of PBS, other buffers such as ammonium bicarbonate, were used in some experiments.. In one embodiment, 10 liters of 0.5M ammonium bicarbonate was added to 1 kg of rice flour.

The resulting homogenate was clarified either by filtration or centrifugation. For the filtration method, the mixture was allowed to settle for about 30 minutes at room temperature, after which the homogenate was collected and filtered. Filters in three different configurations were purchased from Pall Gemansciences and used. They were: a 3 μm pleated capsule, a 1.2 μm serum capsule and a Suporcap capsule 50 (0.2 μm). For centrifugation, a Beckman J2-HC centrifuge was used and the mixture was centrifuged at 30,000 g at 4°C for about 1 hr. The supernatant was retained and the pellet discarded.

In one embodiment, the filtrate and supernatant were further processed, for example by ultra-filtration or dialysis or both to remove components such as lipids, sugars and salt.

The filtrate from the above filtration procedure, which is also called the clarified extract, was then concentrated using a spiral wound tangential flow filter operated in a batch recirculation mode. In one embodiment, PES (polyethersulfone) 3000-4000 molecular weight cutoff membranes were used for this step. These final concentrated extracts were held overnight in a cold room.

The concentrated extracts were next dried to a powder by lyophilization. The lyophilized material was scraped from the lyophilizer trays and combined into a plastic bag. The dry material was compressed by drawing a vacuum on the bag and then the material was blended and the particle size reduced by hand-kneading it through the plastic.

Rice extract can also be produced using a wet milling procedure. Transgenic paddy rice seeds containing recombinant human blood protein can be re-hydrated for a period of 0 to 288 hrs at 30°C. The rehydrated seeds are ground in PBS extraction buffer. The initial seed/buffer ratio can vary over a range such as 1 g/40 ml to 1 g/10 ml.

Over 20% human blood protein can be recovered from the wet milling process. The result of the wet milling becomes an initial extract that may be kept cold (4°C) or stored frozen until use depending on the stability of the blood protein target. The

processing of initial extract to obtain dried extract is the same as that described for dry milling in this section.

EXAMPLE 3

Concentration and diafiltration of recombinant blood protein and control rice extracts.

The conditions used in concentration and diafiltration vary depending on volume, speed, cost, etc. These conditions are standard in the art based on the description herein. The frozen initial extract was thawed in the coldroom (about 2-8°C) for six hours. The thawed material was clarified though a 0.45 μm filter and concentrated using a 5000 Nominal Molecular Weight Cutoff membrane of Polyethersultone.

90 ml of the filtrate of control extract was concentrated to 10 ml and additional 10 ml of deionized water can be added to the concentrated filtrate. The diluted filtrate can be diafiltrated one more time using water. The precipitate starts forming at 16 mS and increases as the ionic strength decreases. A solution of 1.0M ammonium bicarbonate was added to the retentate to add ionic strength. The haze decreases although does not disappear completely. The material was diafiltered multiple times, in one embodiment three times, with water and multiple times, in one embodiment three times, with 0.1 M ammonium bicarbonate. It was concentrated to 9 ml and the membrane is rinsed with 0.1 M ammonium bicarbonate. The concentrate was filtered through several 0.2 μm button filters. In one embodiment, 2.3 ml of the filtrate is lyophilized as is; 2.3 ml of the filtrate is diluted to 12 ml with deionized water and lyophilized, and 2.0 ml of the filtrate is diluted to 25 ml with deionized water and lyophilized. All the filtrates remained clear.

A total of 89 ml of the filtrate of recombinant protein extract was concentrated to 10 ml, and additional 10 ml of 0.1 M ammonium bicarbonate is added. The resulting mixture is concentrated back to 10 ml and another 10 ml of 0.1 M ammonium bicarbonate is added. The retentate starts to haze up. The material was diafiltered multiple times, in one embodiment three times, with 0.1 M ammonium bicarbonate. It was concentrated to 9 ml and the membrane is rinsed with 0.1 M ammonium bicarbonate. The concentrate was filtered through several 0.45 μm button filters. In one embodiment, 2.0 ml of the filtrate was lyophilized as is; 2.0 ml of the filtrate was diluted to 12 ml with deionized water where a haze formed, and lyophilized, and 2.0 ml

of the filtrate was diluted to 12 ml with 0.1 M ammonium bicarbonate that remained clear, and lyophilized.

EXAMPLE 4

Comparison of trial extraction of recombinant protein rice with PBS and ammonium bicarbonate

The conditions used in concentration and diafiltration vary depending on volume, speed, cost, etc. These conditions are all standard in the art based on the description herein. Recombinant protein rice flour is mixed with extraction buffer at about 100 g/L for about 1 hour using a magnetic stir bar. In one 2L beaker, the extraction buffer is PBS, pH 7.4 plus 0.35 M NaCl. In another 2L beaker, the extraction buffer is 0.5 M ammonium bicarbonate. A 15 cm Buchner funnel is pre-coated with about 6g of Cel-pure C300 before adding another 20g of Cel-pure C300. The mixture is filtered at about 3-4 Hg. It is then washed twice with about 100 ml of respective extraction buffer. The extracted filtrate is collected and concentrated with ultra-filtration cartridges: 5K Regenerate Cellulose, 5K PES, and 1K Regenerated Cellulose. The concentrates are lyophilized and analyzed for recombinant blood protein activity contents. The ammonium bicarbonate and PBS, pH7.4 plus 0.35 M NaCl both extract approximately the same amount of rAAT. There is little loss of recombinant protein units in the permeate with any of the ultrafiltration units that were used.

Other extraction buffer can also be used to extract recombinant proteins expressed in transgenic rice grains, for example Tris buffer, ammonium acetate, depending on applications.

EXAMPLE 5

Production of rice extracts containing recombinant blood proteins

The conditions used in concentration and diafiltration vary depending on volume, speed, cost, etc. These conditions are all standard in the art based on the description herein. All equipment is soaked in hot 0.1M NaOH at a starting temperature of about 55°C. Rice flour is added to an about 250-500 gal stainless steel tank containing 0.5M ammonium bicarbonate in a ratio of 95-105 g/L. It is mixed for about 60-80 minutes at about 9°C.

12 plates of 36 inch filter press C300 were pre-coated with about 3-6 kg Cel-pure C300. About 19-26 g/L of Cel-pure is added to the extract and mixed thoroughly. The mixture is pressed at a pressure of about 22 psi at a flow rate of about 82 liters/minute. The filtrate is collected into a 250 gal stainless steel tank and washed with 0.5M ammonium bicarbonate. The press is blown dry. This process is carried out at about 10°C.

The 300 NMW cut-off membranes (Polysulfone), which had been cleaned and stored with 0.1M NaOH after control run is rinsed thoroughly with deionized water. The extract is concentrated and bumped to a 100 gal stainless steel tank. The membrane and the concentration tank were flushed with 0.1M ammonium bicarbonate to recover all remaining extract. The products were covered with plastic and left in the 100 gal tank overnight at room temperature. The concentrate is filtered through spiral wound 1 μm filter and into 5 gal poly container.

EXAMPLE 6

Blending of rice extract containing recombinant proteins into parenteral, inhalant, intranasal and topical formulations.

Recombinant blood proteins (such as AAT) can be highly purified grains from cereal grains for use in medical/pharmaceutical applications. A purification protocol for rice seed extract expressed human AAT has been developed [Huang *et al*, 2002], consisting of preparing a rice seed extract according to the above examples and further purifying the extract preparation using Con-A, DEAE and Octyl Sepharose chromatography respectively. AAT can be purified to greater than 90% homogeneity utilizing such a procedure [Huang *et al*, 2002]. Purified AAT can be utilized in potential pharma/medical applications for the following indications: AAT augmentation/replacement therapy [Sandhaus, 1993; Hubbard *et al*, 1989], cystic fibrosis [McElvaney *et* al,1991; Allen,1996], psoriasis, panniculitis and cutaneous vasculitis [O'Riordan *et al*, 1997; Dowd *et al*, 1995] and pulmonary inflammation [Bingle and Tetley, 1996]. For some of these indications, purified AAT protein preparations can be administered via intravenous (iv) methods in 0.09% saline solution. Alternatively, the saline solution solution could be buffered with serum albumen at 0.5% or some other pharmacologically acceptable protein carrier molecule. AAT dosages are usually around 60 mg/kg. For aerosol delivery, an aerosol generating system can be employed utilizing

a compressed air driven nebulizer selected on the basis of the basis of its ability to generate an aerosol with droplets of the optimum size (< 3um in aerodynamic diameter) for deposition in the lower respiratory tract [Hubbard *et al*, 1989]. Again proteins can either be suspended in sterile water or a buffered saline solution containing a pharmacologically acceptable protein carrier. Alternatively, a dried protein powder containing the purified protein component could be utilized as the dispersal agent and this could be an a rice based extract where the AAT component is greater but not less then 50% by weight.

In another case, recombinant rice expressed and extracted human blood proteins such as AAT and fibrinogen can be employed topically. The use of fibrin sealants/bandages has been a widely accepted used by the medical community. Fibrin sealants are effective hemostatic agents [Mankad and Codispoti, 2001], a means for achieving tissue adhesion, preventing fluid accumulation and promotion of wound healing [Spotnitz, 2001]. Fibrin sealants can also be used as a means of slowly releasing medications, including antibiotics, growth factors and other agents [Spotnitz, 1997]. Rice expressed fibrinogen can also provide a potential low cost and animal virus free source for these indications.

References

1. Allen, E.D. (1996). Opportunities for the use of aerosolized α1-antitrypsin for the treatment of cystic fibrosis. CHEST **110**: 256S-260S.
2. Ausubel F.M. *et al.,* CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y., 1993.
3. Beatty, K., Beath, J. and Travis, J. (1980). Kinetics of association of serine proteinase with native and oxidized α-1-proteinase inhibitor and α-1-antichymotrypsin. Jour Biol Chem **255**: 3931-34.
4. Bednarek, S.Y. and Raikel, N.V. (1992). Intracellular trafficking of secretory proteins. Plant Mol Biol **20**: 133-50.
5. Bhagavan, N.V. (2002). "Biochemistry of Hemostatis" in Medical Biochemistry, Academic Press Inc, Harcourt Inc, Orlando, Fla.
6. Bhagavan, N.V. (2002). "Molecular Immunology" in Medical Biochemistry, Academic Press Inc, Harcourt Inc, Orlando, Fla.

7. Bingle, L. and Tetley, T.D. (1996). Secretory leukoprotease inhibitor; partnering α1-proteinase inhibitor to combat pulmonary inflammation. Thorax **51**:1273-4.

8. Brennan, S.O., Owen, M.C., Boswell, D.R., Lewis, J.H. and Carrell, R.W. (1984). Circulating proalbumin associated with a variant proteinase inhibitor. Biochem Biophys Acta **802**: 24-28.

9. Brown, J.H., Volkmann, N., Jun, G., Henschen-Edman, A.H. and Cohen, C. (2000). The crystal structure of modified bovine fibrinogen. Proc Natl Acad Sci **97**: 85-90.

10. Carrell, R.W., Jeppsson, J.O., Vaughan, L., Brennen, S., Owen, M.C. and Boswell, D.R. (1981). Human α-1-antitrypsin: Carbohydrate attachment and sequence homology. FEBS Lett **135**: 301-3.

11. Carrell, R.W., Jeppson, J.O., Laurrell, C.B., Brennan, S.O., Owen, M.C., Vaughan, L. and Boswell, D.R. (1982). Structure and variation of human alpha-1-antitrypsin. Nature **298**: 329-34.

12. Chrispeels, M.J. (1991). Sorting proteins in the secretory system. Annu Rev Plant Physiol Plant Mol Biol **42**: 21-53.

13. Chrispeels, M.J. and Raikhel, N.V. (1992). Short peptide domains target proteins to plant vacuoles. Cell **68**: 613-16.

14. Christou, P. (1992). Genetic transformation of crop plants using microprojectile bombardment. Plant Jour **2**: 275-281.

15. Daniell, H., Streatfield, S.J. and Wycoff, K. (2001). Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. TRENDS Plant Sci **6**: 219-226.

16. Davidson L.A. and Lonnerdal, B. (1987). Persistence of human milk proteins in the breast-fed infant. Act Pediatric Scand **76**: 733-40.

17. Ditta, G., Stansfield, S., Corbin, D. and Helinski, D.R. (1980). Broad host ramge DNA cloning system from gram-negative bacteria: Construcyionof a gene bank from *Rhizobium meliloti*. Proc Natl Acad Sci USA **77**: 7347-51.

18. Dowd, S.K., Rodgers, G.C. and Callen, J.P. (1995). Effective treqatment with α1-protease inhibitor of chronic cutaneous vasculitis associated with α1-antitrypsin deficiency. Jour Am Acad Dermatol **33**: 913-6.

19. Eriksson, S. (1964). Acta Medic Scand **175**: 197-205.

20. Everse, S.J., Spraggon, G., Veerapandian, L. and Doolittle, R.F. (1999) Conformational changes in fragments D and doubl-D from human fibrinogen upon binding the peptide ligan Gly-His-Pro-amide. Biochemistry **38**: 2941-2946.

21. Friezner-Degen S.J., MacGillivray, R.T.A. and Davie, E.W. (1983). Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin. Biochemistry **22**: 2087-97.

22. Holmberg, N., Gosta, L, Bailey, J.E. and Bulow, L. (1997). Transgenic tobacco expressing *Vitreoscilla* hemoglobin exhibits enhanced growth and altered metabolite production. Nature Biotech **15**: 244-7.

23. Hiatt, A., Cafferkey, R. and Bowdish, K. (1989). Production of antibodies in transgenic plants. Nature **342**: 76-8.

24. Hiatt, A. and Ma, J.K.C. (1992). Monoclonal antibody engineering in plants. FEBS Letters **307**: 71-5.

25. Horvath, H., Huang, J., Wong. O., Kohl E., Okita, T., Kannangara, C.G. and vonWettstein, D. (2000). The production of recombinant proteins in transgenic barley grains. Proc Natl Acad Sci USA **97**: 1914-19.

26. Huang, N., Simmons, C.R. and Rodriguez, R.L. (1990). Codon usage patterns in plant genes. J CAASS **1**: 73-86.

27. Huang, N., Wu, I., Nandi, S., Bowman, E., Huang, J., Sutliff, T. and Rodriguez, R.L. (2001). The tissue-specific activity of a rice B-glucanase promoter (*Gns9*) is used to select rice transformants. Plant Sci **161**: 589-95.

28. Huang, J., Sutliff, T.D., Wu, L., Nandi, S., Benge, K., Terashima, M., Ralston, A.H., Drohan, W., Huang, N. and Rodriguez, R.L. (2001). Expression and purification of functional human α-1-antitrypsin from cultured plant cells. Biotechnol Prog **17**: 126-33.

29. Hubbard, R.C., McElvaney, N.G., Sellers, S.E., Healy, J.T., Czerski, D.B. and Crystal, R.G. (1989). Recombinant DNA-produced α1-antitrypsin administered by aerosol augments lower respiratory tract antineutrophil elastase defenses in individuals with α1-antitrypsin deficiency. Jour Clinical Invest **84**: 1349-54.

30. Jouanin, L., Vilaine, F., d'Enfert, C. and Casse-Delbart, F. (1985). Localization and restriction maps of the replication origin regions of the plasmids of*Agrobacterium rhizogenes* strain A_4. Mol Gen Genet **201**: 370-4.

31. Judah, J.D., Gamble, M. and Steadman, J.H. (1973). Biosynthesis of serum albumin in rat liver, evidence for the exsistance of 'proalbumin'. Biochem J **134**: 1083-91.

32. Kurachi, K., Chandra, T., Degen, S.J., White, T.T., Marchioro, T.L., Woo, S.L. and Davie, E.W. (1981). Cloning and sequence of a cDNA coding for α-1-antitrypsin. Proc Natl Acad Sci **78**: 6826-30.

33. Ma, J.K.C., Hiatt, A., Hein, M., Vine, N.D., Wang, F., Stabila, P., van Dolleweerd, C., Mostov, K. and Lehner, T. (1995). Generation and assembly of Secretory Antibodies in plants. Science **268**: 716-9.

34. Madrazo, J., Bzrown, J.H., Litvinovich, S., Dominguez, R., Yakovlev, S., Medved, L. and Cohen, C. (2001) Crystal structure of the central region of bovine fibrinogen E5 fragment at 1.4 A resolution. Proc Natl Acad Sci **98**: 11967-72.

35. Mankad, P.S. and Codispoti, M. (2001). The role of fibrin sealants in hemostasis. Amer Jour Surg **182**: 21S-28S.

36. McBride, K.E. and Summerfelt, K.R. (1990). Improved binary vectors for *Agrobacterium* mediated plant transformation. Plant Mol Biol **14**: 269-76.

37. McElvaney, N.G., Hubbard, R.C., Birrer, P., Chernick, M.S., Caplan, D.B., Frank, M.M. and Crystal, R.G. (1991). Aerosol α1-antitrypsin treatment for cystic fibrosis. The Lancet **337**: 392-4.

38. McGilligan, K.M., Thomas, D.W. and Eckert, C.D. (1987). Alpha-1-antitrypsin concentration in human milk. Pediatr Res **22**: 268-70.

39. O'Riordan, K., Blei, A., Rao, M.S. and Abecassis, M. (1997). α1-antitrypsin deficiency-associated panniculitis. Transplant **63**: 480-2.

40. Peters, T.J. (1996). All About Albumin: Biochemistry, genetics and medical applications. Academic Press Inc, Harcourt Brace & Company, Orlando, Fla.

41. Rabiet, M.J., Blashill, A., Furie, B. and Furie, B.C. (1986). Prothrombin fragment 1*2*3; a major product of prothrombin activation in human plasma. Jour Biol Chem **261**: 13210-15.

42. Rosenberg, J.S., Beeler, D.L. and Rosenberg, R.D. (1975). Activation of human prothrombin by highly purified human factors V and X_a in the presence of human antithrombin. Jour Biol Chem **250**: 1607-17.

43. Travis, J. and Salvesen, G.S. (1983). Human plasma proteinase inhibitors. Annu Rev Biochem 655-709.

44. Sambrook J. *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (Second Edition), Cold Spring arbor Press, Plainview, N.Y, 1989.

45. Sandhaus, R.A. (1993). "Alpha-1-antitrypsin augmentation therapy" in Proteases, Protease Inhibitors and Protease-Derived Peptides, Birkhauser Verlag, Basel, pp 97-102 Sanford, J., Smith, F.D. and Russell, J.A. (1993). Optimizing the biolistic process for different biological applications. Meth Enzymol **217**: 83-509.

46. Shieh, M.W., Wessler, S.R. and Raikel, N.V. (1993). Nuclear targeting of the maize R protein requires two nuclear localization sequences. Plant Physiol **101**: 353-61.

47. Sijmons, P.C., Dekker, B.M., Schrammeijer, B., Verwoerd, T.C., van den Elzen, P.J.M. and Hoekema, A. (1990). Production of correctly processed human serum albumin in transgenic plants. Bio/Technology **8**: 217-21.

48. Smeekens, S., Bauerle, C., Hageman, J., Keegstra, K. and Weisbeek, P. (1986) The role of the transit peptide in the routing of precursors towards different chloroplast compartments. Cell 46: 365-76.

49. Spotnitz, W.D. (1997). New developments in the use of fibrin sealants: a surgeon's perspective. Jour Long Term Effects Med Implants **7**: 243-53.

50. Spotnitz, W.D. (2001). Commercial fibrin sealants in surgical care. Amer Jour Surg **182**: 8S-14S.Vine, N.D., Drake, P., Hiatt, A. and Ma, J.K.C. (2001). Assembly and plasma membrane targeting of recombinant immunoglobin chains in plants with a murine immunoglobin transmembrane sequence. Plant Molec Biol **45**: 159-67.

51. Tietz, N.W. (1995). Clinical Guide to Laboratory Tests, W.B. Saunders, Philadelphia.

52. Varagona, M.J., Schmidt, R.J. and Raikhel, N.V. (1992). Nuclear localization signals required for nuclear targeting of the maize regulatory protein, Opaque-2. Plant Cell **4**: 1213-27.

53. Vitale, A. and Chrispeels, M.J. (1992). Sorting of proteins to the vacuoles of plant cells. BioEssays **14**: 151-60.

54. Wassmen, C.C., Reiss, B., Bartlett, S.G. and Bohnert, H.J. (1986). The importance of the transit peptides and the transported protein for protein import into chloroplasts. Mol Gen Genet **205**: 446-53.

55. Weeke, B. and Krasilnikoff (1971). A polynomial expression for the serum concentrations of 21 serum proteins from 1 to 93 years of age in normal males and females. Protein Biol Fluids **18**: 173-9.

56. Yakovlev, S., Litvinovich, S., Loukinov, D. and Medved, L. (2000). Role of the beta-strand in the central domain of fibrinogen gamma module. Biochemistry **39**: 15721-9.

57. Yang, Z., Mochalkin, I., Veerapandian, L., Riley, M. and Doolittle, R.F. (2000) Crystal structure of native chicken fibrinogen at 5.5 A resolution. Proc Natl Acad Sci **97**: 3907-12.

All publications cited herein are expressly incorporated herein by reference.

WHAT IS CLAIMED IS:

1. A method of producing a recombinant human blood protein in monocot plant seeds, comprising the steps of:

(a) transforming a monocot plant cell with a chimeric gene comprising

(i) a promoter from the gene of a maturation-specific monocot plant storage protein,

(ii) a first DNA sequence, operably linked to said promoter, encoding a monocot plant seed-specific signal sequence capable of targeting a polypeptide linked thereto to a monocot plant seed endosperm cell, and

(iii) a second DNA sequence, linked in translation frame with the first DNA sequence, encoding a human blood protein, wherein the first DNA sequence and the second DNA sequence together encode a fusion protein comprising an N-terminal signal sequence and the human blood protein;

(b) growing a monocot plant from the transformed monocot plant cell for a time sufficient to produce seeds containing the human blood protein; and

(c) harvesting the seeds from the plant,

wherein the human blood protein constitutes at least about 3.0% of the total soluble protein in the harvested seeds.

2. The method of claim 1, wherein the human blood protein constitutes at least about 5.0% of the total soluble protein in the harvested seeds.

3. The method of claim 1, wherein the human blood protein constitutes at least about 10.0% of the total soluble protein in the harvested seeds.

4. The method of claim 1, further comprising purifying the human blood protein from the harvested seeds.

5. The method of claim 1, wherein the human blood protein is selected from the group consisting of hemoglobin, alpha-1-antitrypsin, fibrinogen, human serum albumin, thrombin, an antibody, and a blood coagulation factor.

6. The method of claim 1, wherein the human blood protein produced in the method comprises one or more plant glycosyl groups.

7. A purified human blood protein obtained by the method of claim 1, wherein the human blood protein comprises one or more plant glycosyl groups.

8. The human blood protein of claim 7, selected from the group consisting of hemoglobin, alpha-1-antitrypsin, fibrinogen, human serum albumin, thrombin, an antibody, and a blood coagulation factor.

9. A monocot plant seed product selected from the group consisting of whole seed, flour, extract and malt, prepared from the harvested seeds obtained by the method of claim 1, wherein the human blood protein constitutes at least about 3.0% of the total soluble protein in the seed product.

10. The seed product of claim 9, wherein the human blood protein constitutes at least about 5.0% of the total soluble protein in the seed product.

11. The seed product of claim 9, wherein the human blood protein constitutes at least about 10.0% of the total soluble protein in the seed product.

12. The seed product of claim 9, wherein the human blood protein is selected from the group consisting of hemoglobin, alpha-1-antitrypsin, fibrinogen, human serum albumin, thrombin, an antibody, and a blood coagulation factor.

13. A composition comprising a purified human blood protein comprising at least one plant glycosyl group and at least one pharmaceutically acceptable excipient or nutrient, wherein the human blood protein is produced in a monocot plant containing a nucleic acid sequence encoding the human blood protein and is purified from seed harvested from the monocot plant, and wherein the at least one nutrient is from a source other than the monocot plant.

14. The composition of claim 13, wherein the human blood protein is selected from the group consisting of hemoglobin, alpha-1-antitrypsin, fibrinogen, human serum albumin, thrombin, an antibody, and a blood coagulation factor.

15. The composition of claim 13, wherein the composition is substantially free of contaminants of the monocot plant.

16. The composition of claim 13, wherein the at least one nutrient is selected from the group consisting of salts, saccharides, vitamins, minerals, amino acids, peptides, and proteins other than the human blood protein.

17. The composition of claim 13, wherein the composition is formulated for parenteral, enteric, inhalation, intranasal or topical delivery.

18. The composition of claim 13, wherein the composition is formulated for enteric delivery, contains at least one pharmaceutically acceptable excipient, and is non-nutritional.

Figure 1. Restriction maps of plasmids (pAPI 398), (pAPI 417) and (pAPI 327); containing the codon-optomized human fibrinogen genes for α, β and γ genes respectively each under the control of the rice glutelin promoter.

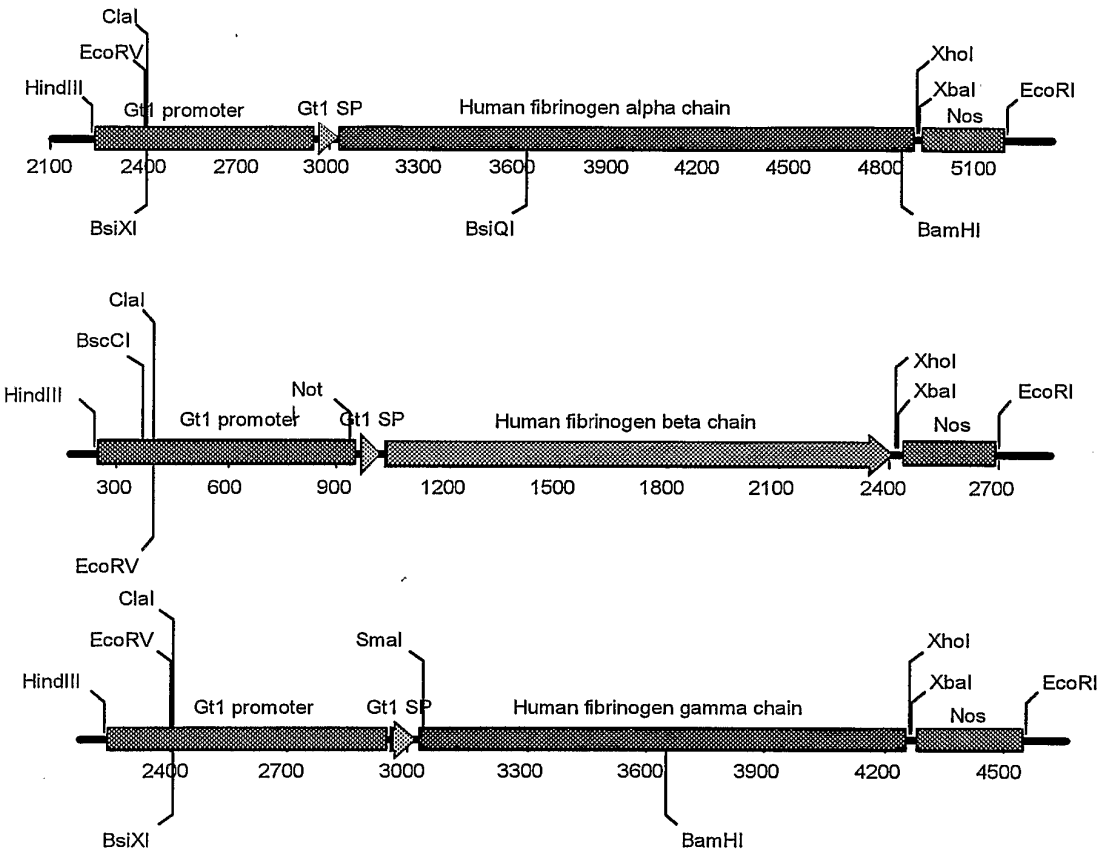


Figure 2. Western blot analysis of independent transgenic rice lines expressing individual subunits of human fibrinogen.

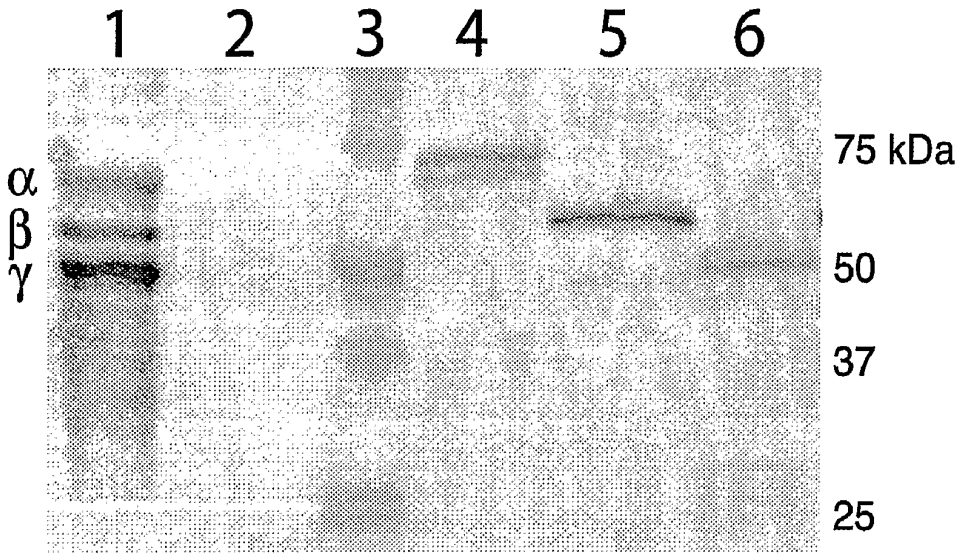


Figure 3. Expression of the fibrinogen polypeptide subunits α,β and γ simultaneously in transgenic rice seeds extracted under non-denaturing and denaturing conditions.

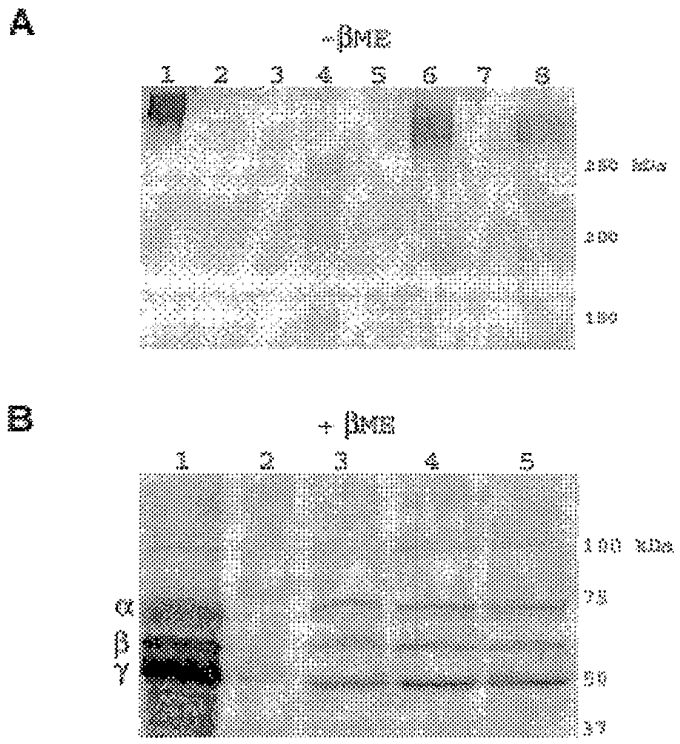


Figure 4. Construct showing the chimeric gene for the expression of alpha-1-antitrypsin in transgenic monocot seeds.

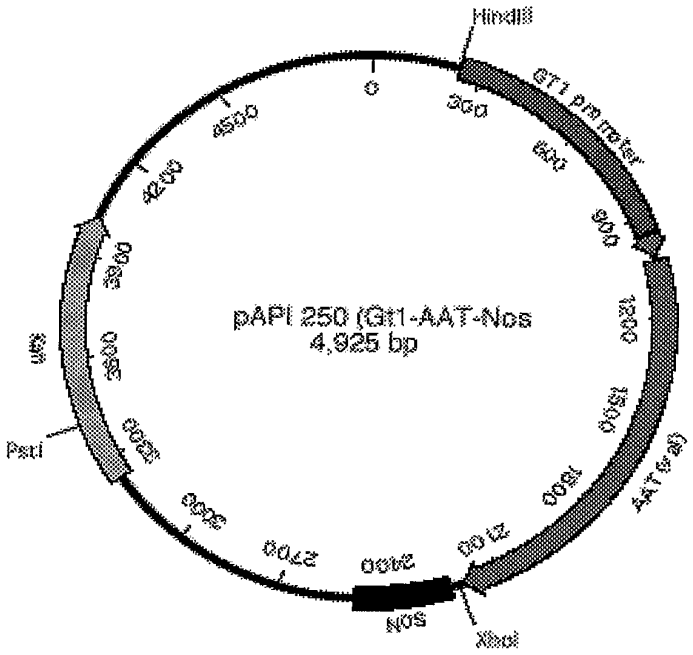


Figure 5. Coomassie-stained gel of total soluble proteins obtained from transgenic rice (var. Kitaake) seed extracts.

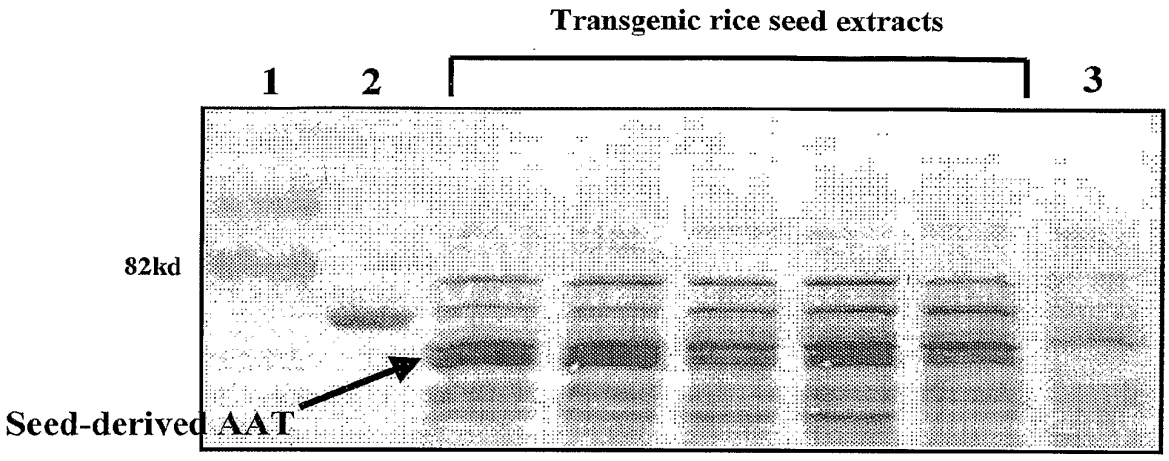


Figure 6. Western blot analysis of recombinant human AAT expressed in transgenic rice grains.

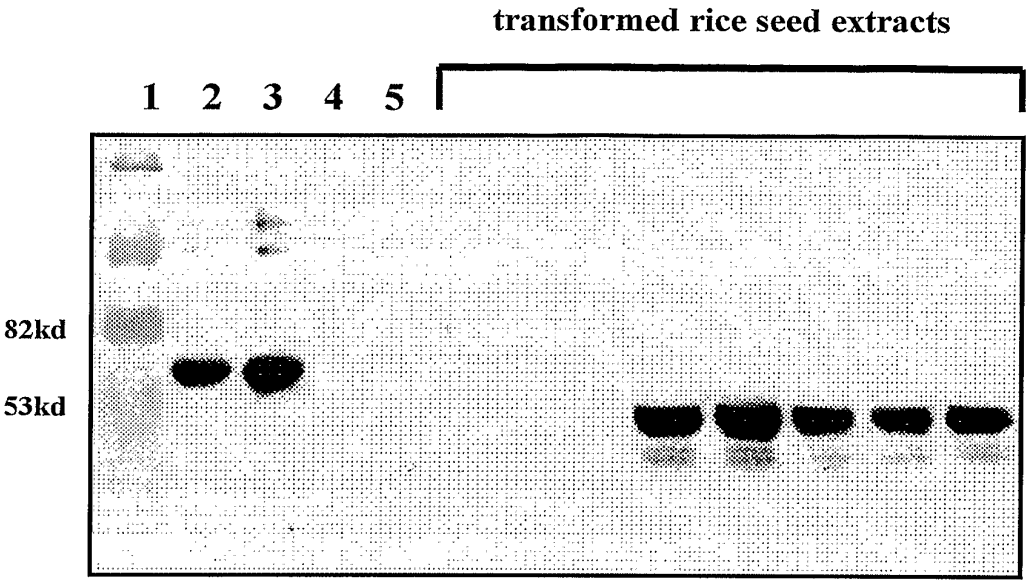


Figure 7. Activity of recombinant AAT against purified porcine elastase (PPE) as determined by Coomassie staining and Western blot analysis.

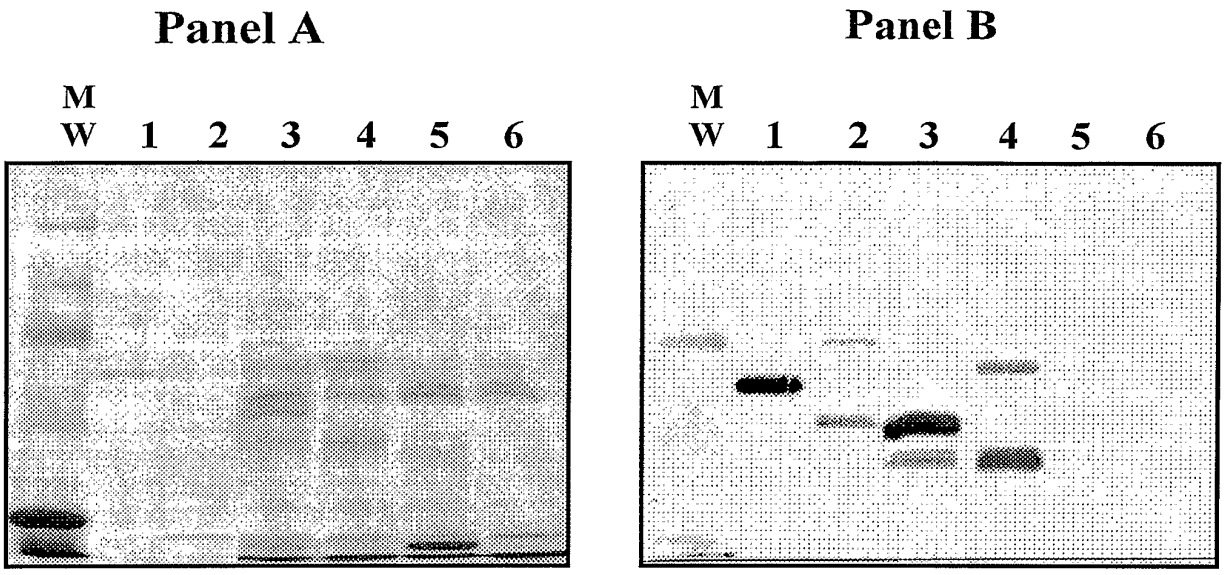


Figure 8. Purification of active AAT from inactive AAT via octyl-sepharose column chromotography.

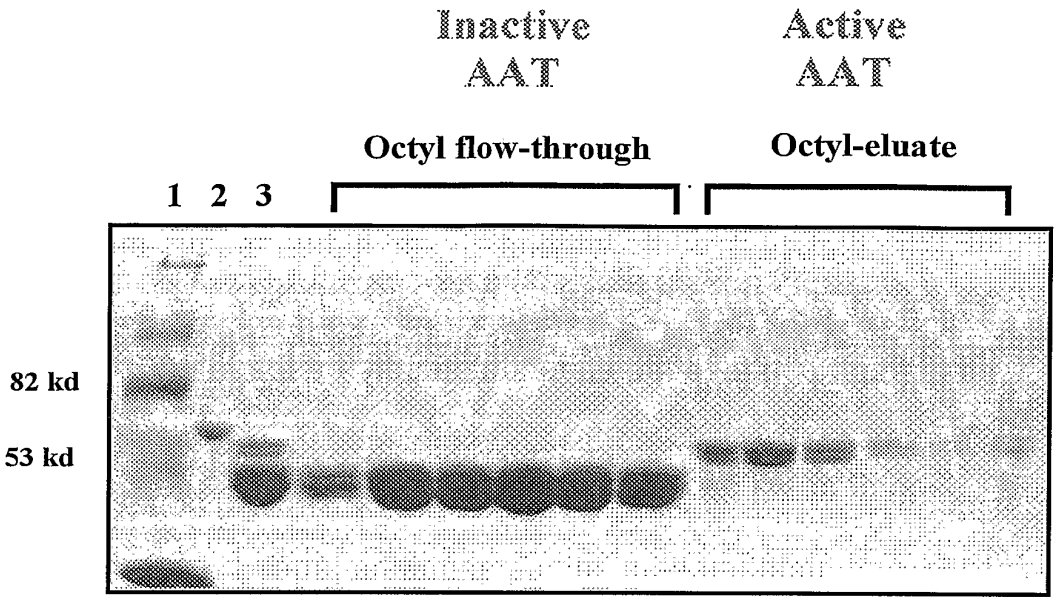


Figure 9. Initial biochemical characterization of AAT purified from rice seed extracts.

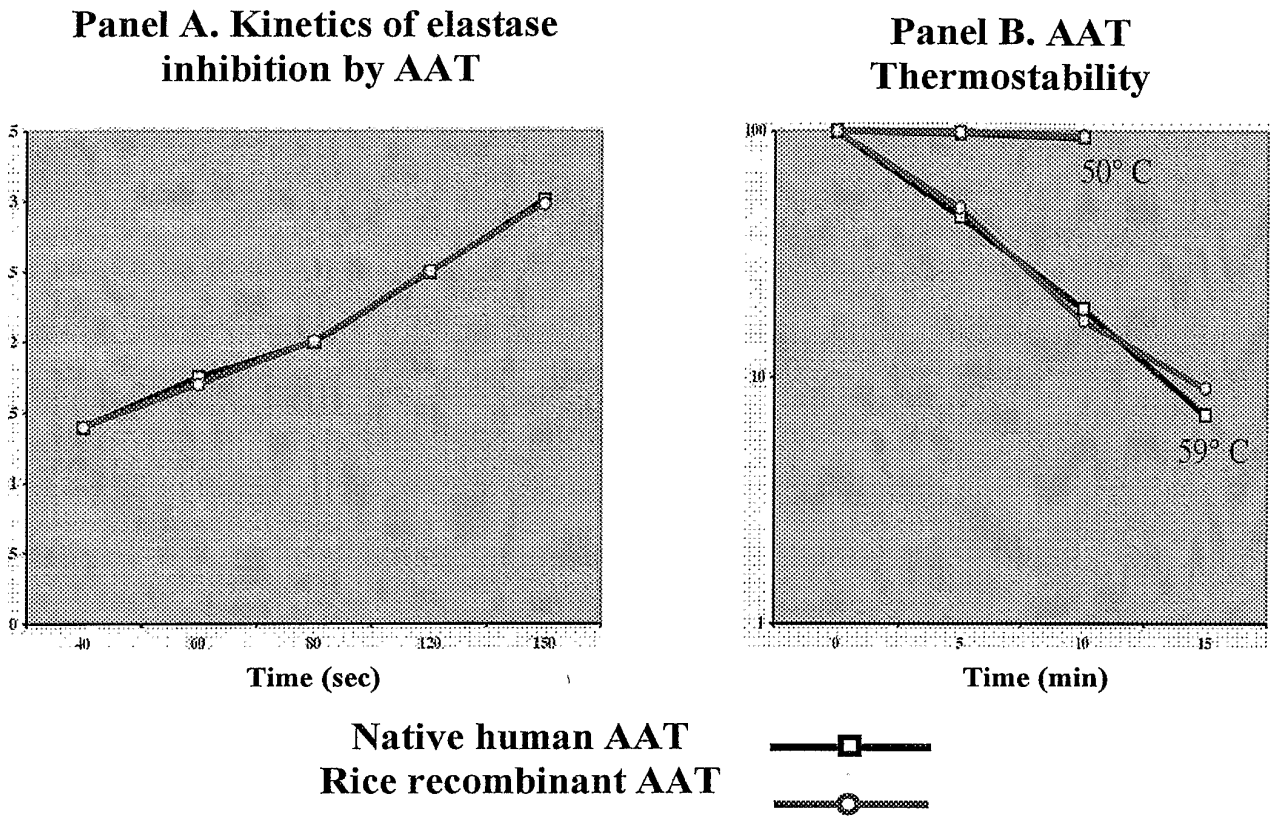


Fig 10. Construct map for the expression of human serum albumen in transgenic rice utilizing the Ramy 1A promoter.

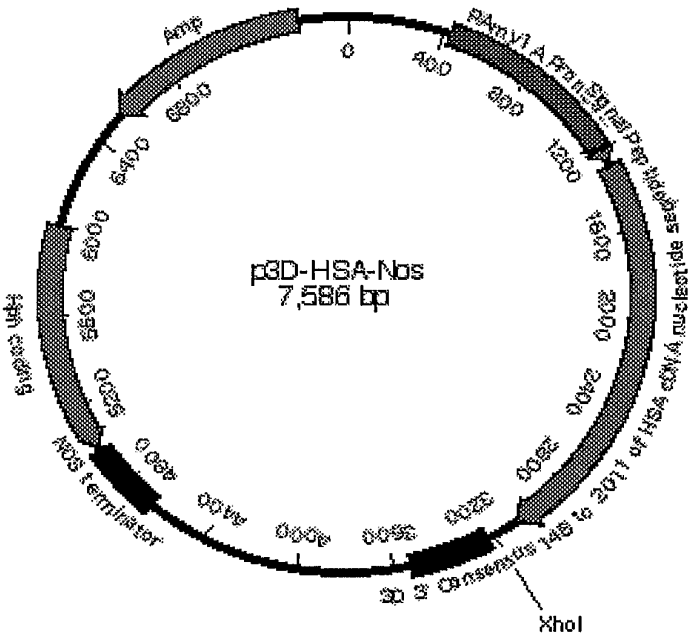
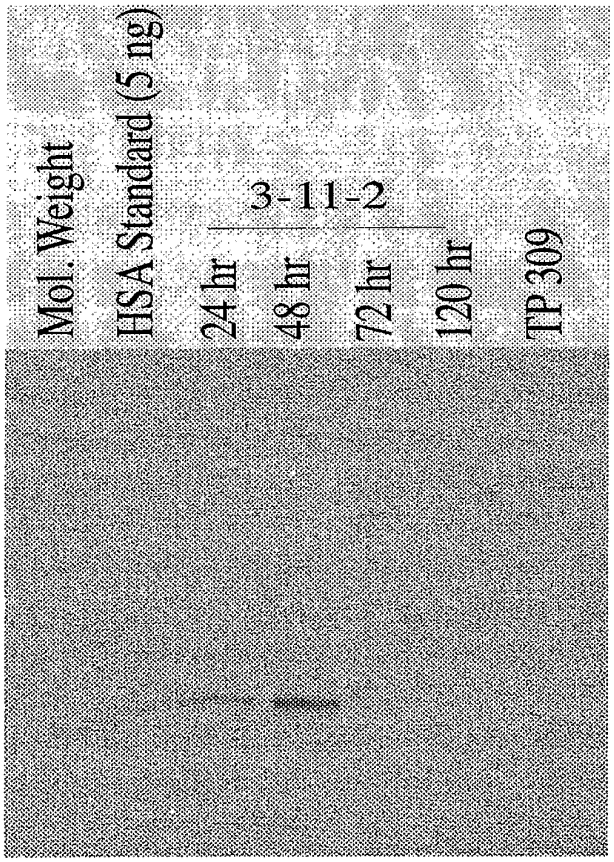


Figure 11. Western blot describing the expression of human serum albumen produced in transgenic rice seeds. Pooled seed from transgenic rice line 3-11-2 were imbibed in water for 24 hours, then 2μM gibberelic acid (GA) added.



SEQUENCE LISTING

<110> HUANG, NING
RODRIQUEZ, RAYMOND L.
HAGIE, FRANK E.
STALKER, DAVID M.

<120> HUMAN BLOOD PROTEINS EXPRESSED IN MONOCOT SEEDS

<130> 023231-00003

<140>
<141>

<150>
<151>

<160> 21

<170> PatentIn Ver. 2.1

<210> 1
<211> 1878
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Codon-optimized
fibrinogen a-polypeptide coding sequence

<400> 1
gccgactccg gcgagggcga cttcctcgcc gagggcggcg gcgtccgggg gccgcgcgtc 60
gtcgagcggc accagtcggc ctgcaaggac tccgactggc cgttctgctc ggacgaggac 120
tggaactaca agtgcccctc cggctgccgc atgaaggggc tgatcgacga ggtcaaccag 180
gacttcacca accgcatcaa caagctcaag aactcgctgt tcgagtacca gaagaacaac 240
aaggactccc actccctcac gaccaacatc atggagatcc tgcgcggcga cttctcctcc 300
gcgaacaacc gcgacaacac ctacaaccgc gtctcggagg acctccgctc ccgcatcgag 360
gtcctgaagc ggaaggtgat cgagaaggtc cagcacatcc agctcctcca aaagaacgtc 420
cgcgcccagc tcgtggacat gaagcgcctg gaggtggaca tcgacatcaa gatccggtcg 480
tgccgcggca gctgctcccg cgccctcgcc cgcgaggtgg acctcaagga ctacgaggac 540
cagcagaagc agctggagca ggtcatcgcc aaggacctcc tcccgagccg cgaccggcag 600
cacctcccac tgatcaagat gaagccggtg ccagacctgg tccccggcaa cttcaagagc 660
cagctccaga aggtcccgcc ggagtggaag gccctcacgg acatgcccca aatgcgcatg 720
gagctggagc gccccggcgg caacgagatc acgcggggcg gctccacctc gtacggcacg 780
ggctccgaga ccgagagccc ccgcaacccc tcctccgccg gctcgtggaa ctcggggtcc 840
agcggccccg gttccacggg caaccgcaac cccggctcgt ccggcaccgg tggcaccgcc 900
acgtggaagc caggtagctc ggggccgggc agcaccggca gctggaactc cggcagcagc 960
ggcaccggct cgacgggcaa ccagaacccg ggcagccccc gccccggctc cacggggacc 1020
tggaacccag gctcctccga gcgggggtcc gccggccact ggacgagcga gagctcggtg 1080
tcgggctcga ccggccagtg gcactcggaa tccggcagct tccggccaga ctcccccggc 1140
agcggcaacg cccggccgaa caacccagac tggggcacct tcgaggaggt ctccggtaac 1200
gtgagccccg gcacgcgccg ggagtaccac acggagaagc tggtgacgtc gaagggcgac 1260
aaggagctcc ggaccggcaa ggagaaggtg acctccggct cgaccaccac cacccggcgg 1320
tcctgctcga agaccgtgac gaagaccgtc atcggtccgg acggccacaa ggaggtcacc 1380
aaggaggtgg tcaccagcga ggacggctcg gactgcccgg aggccatgga cctgggcacc 1440
ctcagcggca tcggcacgct ggacggcttc cgccaccggc acccggacga ggccgccttc 1500
ttcgacaccg ctagcaccgg caagaccttc cccggtttct tctcgccgat gctcggcgag 1560
ttcgtgtccg agacggagag ccggggcagc gagagtggca tcttcaccaa caccaaggaa 1620
tcctcctcgc accacccagg tatcgcggag ttcccgagcc gggggaagtc ctcctcctac 1680

tccaagcagt tcaccagctc cacctcctac aaccggggcg acagcacgtt cgagagcaag 1740
agctacaaga tggccgacga ggccggttcc gaggccgacc acgagggcac ccactcgacg 1800
aagcgcgggc acgccaagtc gcgcccagtg cgcgggatcc acacgtcccc gcgtcgcaag 1860
cccagcctgt ccccgtga 1878

<210> 2
<211> 1386
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Codon-optimized fibrinogen B-polypeptide coding sequence

<400> 2
caaggggtga acgacaacga ggagggtttc ttctccgccc gcggccaccg cccgctggac 60
aagaagcgcg aggaggcccc gagcctgcgc cccgcgcccc cccccatctc cggcggcggc 120
taccgggccc ggccggcaaa ggcggcggca acgcagaaga aggtcgagcg gaaggccccg 180
gacgccggcg gctgcctgca cgcggacccg gacctcggcg tcctgtgccc aacggggtgc 240
cagctccaag aggcgctcct ccaacaggag cgccccatcc ggaactccgt agacgagctg 300
aacaacaacg tggaggcagt gagccagacc tcctcgtcca gcttccagta catgtacctc 360
ctcaaggacc tctggcagaa gcggcaaaag caggtgaagg acaacgagaa cgtggtgaac 420
gagtacagct ccgagctcga aaagcaccag ctgtacatcg acgagaccgt gaactcgaac 480
atccccacga acctccgcgt cctgcgctcg atcctggaga acctccggag caagatccag 540
aagctagaat ccgacgtgtc ggcccagatg gagtattgcc ggaccccgtg caccgtcagc 600
tgcaacatcc cggtggtcag cggcaaggag tgcgaggaga tcatccgcaa gggcggcgag 660
accagcgaga tgtacctcat ccaacccgat tcctccgtca agccataccg ggtgtactgc 720
gacatgaaca cggagaacgg cgggtggacc gtgatccaga accgccagga cggctccgtg 780
gacttcggcc gcaagtggga cccgtacaag cagggcttcg gcaacgtggc cacgaacacg 840
gacgggaaga actactgcgg gctccccggc gaatactggc tgggcaacga caagatctcc 900
cagctgaccc gcatgggccc caccgagctg ctcatcgaga tggaggactg gaagggcgac 960
aaggtgaagg cccactacgg gggcttcacg gtgcagaacg aggcgaacaa gtaccaaatc 1020
tcggtgaaca agtaccgcgg caccgctggg aacgcgctca tggacggcgc gagccagctg 1080
atgggcgaga accgcaccat gaccatccac aacggcatgt tcttcagcac ctacgaccgc 1140
gacaacgacg ggtggctcac gagcgacccc cggaagcagt gctcgaagga ggacggcggc 1200
ggctggtggt acaaccgctg ccacgcggca aaccccaacg gtcgctacta ctggggcggt 1260
cagtacacgt gggacatggc gaagcacggc accgacgacg gcgtcgtctg gatgaactgg 1320
aagggctcgt ggtacagcat gcggaagatg tccatgaaga tccgcccctt cttccccag 1380
cagtga 1386

<210> 3
<211> 1236
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Codon-optimized fibrinogen y-polypeptide encoding sequence

<400> 3
tacgtcgcca cccgggacaa ctgctgcatc ctggacgagc ggttcgggag ctactgccca 60
accacctgcg gcatcgccga cttcctgtcc acgtaccaga cgaaggtgga caaggacctc 120
cagtccctgg aggacatcct ccaccaggtg gagaacaaga cgtcggaggt caagcagctc 180
atcaaggcca tccagctcac ctacaacccg gacgaatcgt ccaagcccaa catgatcgac 240
gccgccaccc tcaagtcgcg gaagatgctg gaggagatca tgaagtacga ggcgtccatc 300
ctcacccacg actcctccat ccgctacctc caggagatct acaactccaa caaccaaaag 360
atcgtcaacc tcaaggagaa ggtcgcccag ctggaggcgc aatgccagga gccctgcaag 420

gacacggtgc aaatccacga catcacgggg aaggactgcc aagacatcgc caacaagggc 480
gccaagcaga gcgggctcta cttcatcaag cccctcaagg cgaaccagca gttcctggtc 540
tactgcgaga tcgacggctc gggcaacggc tggaccgtct tccagaagcg cctcgacggc 600
tccgtggact tcaagaagaa ctggatccaa tacaaggagg gcttcggcca cctctccccc 660
accggcacga cggagttctg gctgggcaac gagaagatcc acctcatctc cacgcagagc 720
gcgatcccat acgccctccg ggtggagctg gaggactgga acggccgcac cagcaccgcg 780
gactacgcaa tgttcaaggt gggcccagag gcggacaagt accggctgac ctacgcctac 840
ttcgcgggcg gggacgcggg ggacgccttc gacgggttcg acttcggtga cgacccaagc 900
gacaagttct tcacgtccca caacggtatg cagttcagca cgtgggacaa cgacaacgac 960
aagttcgagg gtaactgcgc ggagcaggac ggcagcggct ggtggatgaa caagtgccac 1020
gcgggccacc tcaacggcgt ctactaccag ggcgggacct acagcaaggc atccacgcca 1080
aacgggtacg acaacggtat catctgggcc acgtggaaga cgcgctggta cagcatgaag 1140
aagaccacca tgaagatcat cccgttcaac cggctgacca tcggtgaggg ccagcagcac 1200
cacctcggcg gggccaagca ggcgggcgac gtgtga 1236

<210> 4
<211> 1748
<212> DNA
<213> Homo sapiens

<400> 4
gatgcacaca agagtgaggt tgctcatcgg tttaaagatt tgggagaaga aaatttcaaa 60
gccttggtgt tgattgcctt tgctcagtat cttcagcagt gtccatttga agatcatgta 120
aaattagtga atgaagtaac tgaatttgca aaaacatgtg ttgctgatga gtcagctgaa 180
aattgtgaca aatcacttca tacccttttt ggagacaaat tatgcacagt tgcaactctt 240
cgtgaaacct atggtgaaat ggctgactgc tgtgcaaaac aagaacctga gagaaatgaa 300
tgcttcttgc aacacaaaga tgacaaccca aacctccccc gattggtgag accagaggtt 360
gatgtgatgt gcactgcttt tcatgacaat gaagagacat ttttgaaaaa atacttatat 420
gaaattgcca gaagacatcc ttacttttat gccccggaac tccttttctt tgctaaaagg 480
tataaagctg cttttacaga atgttgccaa gctgctgata aagctgcctg cctgttgcca 540
aagctcgatg aacttcggga tgaagggaag gcttcgtctg ccaaacagag actcaagtgt 600
gccagtctcc aaaaatttgg agaaagagct ttcaaagcat gggcagtagc tcgcctgagc 660
cagagatttc ccaaagctga gtttgcagaa gtttccaagt tagtgacaga tcttaccaaa 720
gtccacacgg aatgctgcca tggagatctg cttgaatgtg ctgatgacag ggcggacctt 780
gccaagtata tctgtgaaaa tcaagattcg atctccagta aactgaagga atgctgtgaa 840
aaacctctgt tggaaaaatc ccactgcatt gccgaagtgg aaaatgatga gatgcctgct 900
gacttgcctt cattagctgc tgattttgtt gaaagtaagg atgtttgcaa aaactatgct 960
gaggcaaagg atgtcttcct gggcatgttt ttttatgaat atgcaagaag gcatcctgat 1020
tactctgtcg tgctgctgct gagacttgcc aagacatatg aaaccactct agagaagtgc 1080
tgtgccgctg cagatcctca tgaatgctat gccaaagtgt tcgatgaatt taaacctctt 1140
gtggaagagc ctcagaattt aatcaaacaa aattgtgagc tttttgagca gcttggagag 1200
tacaaattcc agaatgcgct attagttcgt tacaccaaga aagtacccca agtgtcaact 1260
ccaactcttg tagaggtctc aagaaaccta ggaaaagtgg gcagcaaatg ttgtaaacat 1320
cctgaagcaa aaagaatgcc ctgtgcagaa gactatctat ccgtggtcct gaaccagtta 1380
tgtgtgttgc atgagaaaac gccagtaagt gacagagtca ccaaatgctg cacagaatcc 1440
ttggtgaaca ggcgaccatg cttttcagct ctggaagtcg atgaaacata cgttcccaaa 1500
gagtttaatg ctgaaacatt caccttccat gcagatatat gcacactttc tgagaaggag 1560
agacaaatca agaaacaaac tgcacttgtt gagcttgtga caaggcaaca aaagagcaac 1620
tgaaagctgt tatggatgat ttcgcagctt ttgtagagaa gtgctgcaag gctgacgata 1680
aggagacctg ctttgccgag gagggtaaaa aacttgttgc tgcaagtcaa gctgccttag 1740
gcttataa 1748

<210> 5
<211> 1185
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Codon optimized alpha-1-antitrypsin coding sequence

<400> 5
gaggacccgc agggcgacgc cgcccagaag accgacacca gccaccacga ccaggaccac 60
ccgacgttca acaagatcac cccgaatttg gccgaattcg ccttcagcct gtaccgccag 120
ctcgcgcacc agtccaactc caccaacatc ttcttcagcc cggtgagcat cgccaccgcc 180
ttcgccatgc tgtccctggg taccaaggcg gacacccacg acgagatcct cgaagggctg 240
aacttcaacc tgacggagat cccggaggcg cagatccacg agggcttcca ggagctgctc 300
aggacgctca accagccgga ctcccagctc cagctcacca ccggcaacgg gctcttcctg 360
tccgagggcc tcaagctcgt cgataagttc ctggaggacg tgaagaagct ctaccactcc 420
gaggcgttca ccgtcaactt cggggacacc gaggaggcca agaagcagat caacgactac 480
gtcgagaagg ggacccaggg caagatcgtg gacctggtca aggaattgga cagggacacc 540
gtcttcgcgc tcgtcaacta catcttcttc aagggcaagt gggagcgccc gttcgaggtg 600
aaggacaccg aggaggagga cttccacgtc gaccaggtca ccaccgtcaa ggtcccgatg 660
atgaagaggc tcggcatgtt caacatccag cactgcaaga agctctccag ctgggtgctc 720
ctcatgaagt acctggggaa cgccaccgcc atcttcttcc tgccggacga gggcaagctc 780
cagcacctgg agaacgagct gacgcacgac atcatcacga agttcctgga gaacgaggac 840
aggcgctccg ctagcctcca cctcccgaag ctgagcatca ccggcacgta cgacctgaag 900
agcgtgctgg gccagctggg catcacgaag gtcttcagca acggcgcgga cctctccggc 960
gtgacggagg aggccccccc gaagctctcc aaggccgtgc acaaggcggt gctcacgatc 1020
gacgagaagg ggacggaagc tgccggggcc atgttcctgg aggccatccc cgtgtccatc 1080
ccgcccgagg tcaagttcaa caagcccttc gtcttcctga tgatcgagca gaacacgaag 1140
agccccctct tcatggggaa ggtcgtcaac cccacgcaga agtga 1185

<210> 6
<211> 786
<212> DNA
<213> Oryza sp.

<400> 6
catgagtaat gtgtgagcat tatgggacca cgaaataaaa agaacatttt gatgagtcgt 60
gtatcctcga tgagcctcaa aagttctctc accccggata agaaaccctt aagcaatgtg 120
caaagtttgc attctccact gacataatgc aaaataagat atcatcgatg acatagcaac 180
tcatgcatca tatcatgcct ctctcaacct attcattcct actcatctac ataagtatct 240
tcagctaaat gttagaacat aaacccataa gtcacgtttg atgagtatta ggcgtgacac 300
atgacaaatc acagactcaa gcaagataaa gcaaaatgat gtgtacataa aactccagag 360
ctatatgtca tattgcaaaa agaggagagc ttataagaca aggcatgact cacaaaaatt 420
cacttgcctt tcgtgtcaaa aagaggaggg ctttacatta tccatgtcat attgcaaaag 480
aaagagagaa agaacaacac aatgctgcgt caattataca tatctgtatg tccatcatta 540
ttcatccacc tttcgtgtac cacacttcat atatcataag agtcacttca cgtctggaca 600
ttaacaaact ctatcttaac atttagatgc aagagccttt atctcactat aaatgcacga 660
tgatttctca ttgtttctca caaaaagcgg ccgcttcatt agtcctacaa caacatggca 720
tccataaatc gccccatagt tttcttcaca gtttgcttgt tcctcttgtg cgatggctcc 780
ctagcc 786

<210> 7
<211> 1055
<212> DNA
<213> Oryza sp.

<400> 7
ctgcagggag gagaggggag agatggtgag agaggaggaa gaagaggagg ggtgacaatg 60
atatgtgggg catgtgggca cccaattttt taattcattc ttttgttgaa actgacatgt 120
gggtcccatg agatttatta tttttcggat cgaatcgcca cgtaagcgct acgtcaatgc 180
tacgtcagat gaagaccgag tcaaattagc cacgtaagcg ccacgtcagc caaaaccacc 240

atccaaaccg ccgagggacc tcatctgcac tggttttgat agttgaggga cccgttgtat 300
ctggtttttc gattgaagga cgaaaatcaa atttgttgac aagttaaggg accttaaatg 360
aacttattcc atttcaaaat attctgtgag ccatatatac cgtgggcttc caatcctcct 420
caaattaaag ggccttttta aaatagataa ttgccttctt tcagtcaccc ataaaagtac 480
aaaactacta ccaacaagca acatgcgcag ttacacacat tttctgcaca tttccgccac 540
gtcacaaaga gctaagagtt atccctagga caatctcatt agtgtagata catccattaa 600
tcttttatca gaggcaaacg taaagccgct ctttatgaca aaaataggtg acacaaaagt 660
gttatctgcc acatacataa cttcagaaat tacccaacac caagagaaaa ataaaaaaaa 720
atctttttgc aagctccaaa tcttggaaac ctttttcact ctttgcagca ttgtactctt 780
gctctttttc caaccgatcc atgtcaccct caagcttcta cttgatctac acgaagctca 840
ccgtgcacac aaccatggcc acaaaaaccc tataaaaccc catccgatcg ccatcatctc 900
atcatcagtt cattaccaac aaacaaaaga ggaaaaaaaa catatacact tctagtgatt 960
gtctgattga tcatcaatct agaggcggcc gcatggctag caaggtcgtc ttcttcgcgg 1020
cggcgctcat ggcggccatg gtggccatct ccggc 1055

<210> 8
<211> 976
<212> DNA
<213> Triticum sp.

<400> 8
ctgcaggcca gggaaagaca atggacatgc aaagaggtag gggcagggaa gaaacacttg 60
gagatcatag aagaacataa gaggttaaac ataggagggc ataatggaca attaaatcta 120
cattaattga actcatttgg gaagtaaaca aaatccatat tctggtgtaa atcaaactat 180
ttgacgcgga tttactaaga tcctatgtta attttagaca tgactggcca aaggtttcag 240
ttagttcatt tgtcacggaa aggtgttttc ataagtccaa aactctacca actttttttgc 300
acgtcatagc atagatagat gttgtgagtc attggataga tattgtgagt cagcatggat 360
ttgtgttgcc tggaaatcca actaaatgac aagcaacaaa acctgaaatg ggctttagga 420
gagatggttt atcaatttac atgttccatg caggctacct tccactactc gacatggtta 480
gaagttttga gtgccgcata tttgcggaag caatggcact actcgacatg gttagaagtt 540
ttgagtgccg catatttgcg gaagcaatgg ctaacagata catattctgc caaaccccaa 600
gaaggataat cactcctctt agataaaaag aacagaccaa tgtacaaaca tccacacttc 660
tgcaaacaat acaccagaac taggattaag cccattacgt ggctttagca gaccgtccaa 720
aaatctgttt tgcaagcacc aattgctcct tacttatcca gcttcttttg tgttggcaaa 780
ctgccctttt ccaaccgatt ttgtttcttc tcacgctttc ttcataggct aaactaacct 840
cggcgtgcac acaaccatgt cctgaacctt cacctcgtcc ctataaaagc ccatccaacc 900
ttacaatctc atcatcaccc acaacaccga gcaccccaat ctacagatca attcactgac 960
agttcactga tctaga 976

<210> 9
<211> 1009
<212> DNA
<213> Oryza sp.

<400> 9
ctgcagtaat ggatacctag tagcaagcta gcttaaacaa atctaaattc caatctgttc 60
gtaaacgttt tctcgatcgc aattttgatc aaaactattg aaaacctcaa ttaaaccatt 120
caaaattttt aatataccca acaagagcgt ccaaaccaaa tatgtaaata tggatgtcat 180
gataattgac ttatgacaat gtgattattt catcaagtct ttaaatcatt aattctagtt 240
gaaggtttat gttttcttat gctaaagggt tatgtttata taagaatatt aaagagcaaa 300
ttgcaataga tcaacacaac aaatttgaat gtttccagat gtgtaaaaat atccaaatta 360
attgttttaa aatagtttta agaaggatct gatatgcaag tttgatagtt agtaaactgc 420
aaaagggctt attacatgga aaattcctta ttgaatatgt ttcattgact ggtttatttt 480
acatgacaac aaagttacta gtatgtcaat aaaaaaatac aaggttactt gtcaattgta 540
ttgtgccaag taaagatgac aacaaacata caaatttatt tgttctttta tagaaacacc 600
taacttatca aggatagttg gccacgcaaa aatgacaaca tactttacaa ttgtatcatc 660
ataaagatct tatcaagtat aagaacttta tggtgacata aaaaataatc acaagggcaa 720

gacacatact aaaagtatgg acagaaattt cttaacaaac tccatttgtt ttgtatccaa 780
aagcataaga aatgagtcat ggctgagtca tgatatgtag ttcaatcttg caaaattgcc 840
tttttgttaa gtattgtttt aacactacaa gtcacatatt gtctatactt gcaacaaaca 900
ctattaccgt gtatcccaag tggccttttc attgctatat aaactagctt gatcggtctt 960
tcaactcaca tcaattagct taagtttcca ttagcaactg ctaatagct 1009

<210> 10
<211> 839
<212> DNA
<213> Oryza sp.

<400> 10
ctgcagtgta agtgtagctt cttatagctt agtgctttac tatcttcaca agcacatgct 60
atagtattgt tccaagatga aagaataatt catccttgct accaacttgc atgatattat 120
atttgtgaat atcctatctc ttggcttata atgaaatgtg ctgctgggtt attctgacca 180
tggtatttga gagcctttgt atagctgaaa ccaacgtata tcgagcatgg aacagagaac 240
aaaatgcaag gatttttta ttctggttca tgccctggat gggttaatat cgtgatcatc 300
aaaaaagata tgcataaaat taaagtaata aatttgctca taagaaacca aaaccaaaag 360
cacatatgtc ctaaacaaac tgcattttgt ttgtcatgta gcaatacaag agataatata 420
tgacgtggtt atgacttatt cacttttgt gactccaaaa tgtagtaggt ctaactgatt 480
gtttaaagtg atgtcttact gtagaagttt catcccaaaa gcaatcacta aagcaacaca 540
cacgtatagt ccaccttcac gtaattcttt gtggaagata acaagaaggc tcactgaaaa 600
ataaaagcaa agaaaaggat atcaaacaga ccattgtgca tcccattgat ccttgtatgt 660
ctatttatct atcctccttt tgtgtacctt acttctatct agtgagtcac ttcatatgtg 720
gacattaaca aactctatct taacatctag tcgatcacta ctttacttca ctataaaagg 780
accaacatat atcatccatt tctcacaaaa gcattgagtt cagtcccaca aaatctaga 839

<210> 11
<211> 1302
<212> DNA
<213> Oryza sp.

<400> 11
ctgcagagat atggattttc taagattaat tgattctctg tctaaagaaa aaaagtatta 60
ttgaattaaa tggaaaaaga aaaaggaaaa aggggatggc ttctgctttt tgggctgaag 120
gcggcgtgtg gccagcgtgc tgcgtgcgga cagcgagcga acacacgacg gagcagctac 180
gacgaacggg ggaccgagtg gaccggacga ggatgtggcc taggacgagt gcacaaggct 240
agtggactcg gtccccgcgc ggtatcccga gtggtccact gtctgcaaac acgattcaca 300
tagagcgggc agacgcggga gccgtcctag gtgcaccgga agcaaatccg tcgcctgggt 360
ggatttgagt gacacggccc acgtgtagcc tcacagctct ccgtggtcag atgtgtaaaa 420
ttatcataat atgtgttttt caaatagtta aataatatat ataggcaagt tatatgggtc 480
aataagcagt aaaaaggctt atgacatggt aaaattactt acaccaatat gccttactgt 540
ctgatatatt ttacatgaca acaaagttac aagtacgtca tttaaaaata caagttactt 600
atcaattgta gtgtatcaag taaatgacaa caaacctaca aatttgctat tttgaaggaa 660
cacttaaaaa aatcaatagg caagttatat agtcaataaa ctgcaagaag gcttatgaca 720
tggaaaaatt acatacacca atatgcttta ttgtccggta tattttacaa gacaacaaag 780
ttataagtat gtcatttaaa aatacaagtt acttatcaat tgtcaagtaa atgaaaacaa 840
acctacaaat ttgttatttt gaaggaacac ctaaattatc aaatatagct tgctacgcaa 900
aatgacaaca tgcttacaag ttattatcat cttaaagtta gactcatctt ctcaagcata 960
agagctttat ggtgcaaaaa caaatataat gacaaggcaa agatacatac atattaagag 1020
tatggacaga catttcttta acaaactcca tttgtattac tccaaaagca ccagaagttt 1080
gtcatggctg agtcatgaaa tgtatagttc aatcttgcaa agttgccttt ccttttgtac 1140
tgtgttttaa cactacaagc catatattgt ctgtacgtgc aacaaactat atcaccatgt 1200
atcccaagat gctttttat tgctatataa actagcttgg tctgtctttg aactcacatc 1260
aattagctta agtttccata agcaagtaca aatagctcta ga 1302

<210> 12
<211> 675
<212> DNA
<213> Oryza sp.

<400> 12
ctgcagcatc ggcttaggtg tagcaacacg actttattat tattattatt attattatta 60
ttattttaca aaaatataaa atagatcagt ccctcaccac aagtagagca agttggtgag 120
ttattgtaaa gttctacaaa gctaatttaa aagttattgc attaacttat ttcatattac 180
aaacaagagt gtcaatggaa caatgaaaac catatgacat actataattt tgtttttatt 240
attgaaatta tataattcaa agagaataaa tccacatagc cgtaaagttc tacatgtggt 300
gcattaccaa aatatatata gcttacaaaa catgacaagc ttagtttgaa aaattgcaat 360
ccttatcaca ttgacacata aagtgagtga tgagtcataa tattattttt cttgctaccc 420
atcatgtata tatgatagcc acaaagttac tttgatgatg atatcaaaga acatttttag 480
gtgcacctaa cagaatatcc aaataatatg actcacttag atcataatag agcatcaagt 540
aaaactaaca ctctaaagca accgatggga aagcatctat aaatagacaa gcacaatgaa 600
aatcctcatc atccttcacc acaattcaaa tattatagtt gaagcatagt agtagaatcc 660
aacaacaatc tagag 675

<210> 13
<211> 1098
<212> DNA
<213> Oryza sp.

<400> 13
ccaggcttca tcctaaccat tacaggcaag atgttgtatg aagaagggcg aacatgcaga 60
ttgttaaact gacacgtgat ggacaagaat gaccgattgg tgaccggtct gacaatggtc 120
atgtcgtcag cagacagcca tctcccacgt cgcgcctgct tccggtgaaa gtggaggtag 180
gtatgggccg tcccgtcaga aggtgattcg gatggcagcg atacaaatct ccgtccatta 240
atgaagagaa gtcaagttga aagaaaggga gggagagatg gtgcatgtgg gatccccttg 300
ggatataaaa ggaggacctt gcccacttag aaaggagagg agaaagcaat cccagaagaa 360
tcgggggctg actggcactt tgtagcttct tcatacgcga atccaccaaa acacaggagt 420
agggtattac gcttctcagc ggcccgaacc tgtatacatc gcccgtgtct tgtgtgtttc 480
cgctcttgcg aaccttccac agattgggag cttagaacct cacccagggc ccccggccga 540
actggcaaag gggggcctgc gcggtctccc ggtgaggagc cccacgctcc gtcagttcta 600
aattacccga tgagaaaggg agggggggggg gggaaatctg ccttgtttat ttacgatcca 660
acggatttgg tcgacaccga tgaggtgtct taccagttac cacgagctag attatagtac 720
taattacttg aggattcggt tcctaatttt ttacccgatc gacttcgcca tggaaaattt 780
tttattcggg ggagaatatc caccctgttt cgctcctaat taagatagga attgttacga 840
ttagcaacct aattcagatc agaattgtta gttagcggcg ttggatccct cacctcatcc 900
catcccaatt cccaaaccca aactcctctt ccagtcgccg acccaaacac gcatccgccg 960
cctataaatc ccacccgcat cgagcctatc aagcccaaaa aaccacaaac caaacgaaga 1020
aggaaaaaaa aaggaggaaa agaaaagagg aggaaagcga agaggttgga gagagacgct 1080
cgtctccacg tcgccgcc 1098

<210> 14
<211> 432
<212> DNA
<213> Hordeum sp.

<400> 14
cttcgagtgc ccgccgattt gccagcaatg gctaacagac acatattctg ccaaaacccc 60
agaacaataa tcacttctcg tagatgaaga gaacagacca agatacaaac gtccacgctt 120
cagcaaacag taccccagaa ctaggattaa gccgattacg cggctttagc agaccgtcca 180
aaaaaactgt tttgcaaagc tccaattcct ccttgcttat ccaatttctt ttgtgttggc 240
aaactgcact tgtccaaccg attttgttct tcccgtgttt cttcttaggc taactaacac 300
agccgtgcac atagccatgg tccggaatct tcacctcgtc cctataaaag cccagccaat 360

ctccacaatc tcatcatcac cgagaacacc gagaaccaca aaactagaga tcaattcatt 420
gacagtccac cg 432

<210> 15
<211> 60
<212> DNA
<213> Triticum sp.

<400> 15
atggctaagc gcctggtcct ctttgcggca gtagtcgtcg ccctcgtggc tctcaccgcc 60

<210> 16
<211> 72
<212> DNA
<213> Oryza sp.

<400> 16
atggcaacta ccatttctc tcgtttttct atatactttt gtgctatgct attatgccag 60
ggttctatgg cc 72

<210> 17
<211> 85
<212> DNA
<213> Oryza sp.

<400> 17
atgtggacat taacaaactc tatcttaaca tctagtcgat cactacttta cttcactata 60
aaaggaccaa catatatcat ccatt 85

<210> 18
<211> 72
<212> DNA
<213> Oryza sp.

<400> 18
atggcgagtt ccgttttctc tcggttttct atatactttt gtgttcttct attatgccat 60
ggttctatgg cc 72

<210> 19
<211> 69
<212> DNA
<213> Oryza sp.

<400> 19
atgaagatca ttttcgtatt tgctctcctt gctattgttg catgcaacgc ttctgcacgg 60
tttgatgct 69

<210> 20
<211> 63
<212> DNA
<213> Oryza sp.

<400> 20

atggccgccc gcgccgccgc cgccgcgttc ctgctgctgc tcatcgtcgt tggtcaccgc 60
gcc 63

<210> 21
<211> 63
<212> DNA
<213> Hordeum sp.

<400> 21
atggctaagc ggctggtcct ctttgtggcg gtaatcgtcg ccctcgtggc tctcaccacc 60
gcc 63

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/09209

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12N 15/00, 15/63, A01H 5/00
US CL : 536/23.1, 23.5, 23.52, 24.1; 435/320.1, 69.6; 800/298, 320, 320.2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.1, 23.5, 23.52, 24.1; 435/320.1, 69.6; 800/298, 320, 320.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST, STN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,888,789 A (RODRIGUEZ) 30 March 1999 (30-03-1999), column C11, lines 42-69.	1-18
Y	US 5,994,628 A (RODRIGUEZ) 30 November 1999 (30-11-1999), column C11, lines 36-54.	1-18
Y	US 5,693,506 A (RODRIGUEZ) 2 December 1997 (02-12-1997), column 8, lines 51 to col 9, line 12.	1-18
Y	US 5,889,189 A (RODRIGUEZ) 30 March 1999 (30-03-1999), column 11, lines 42-65.	1-18

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
26 July 2003 (26.07.2003)

Date of mailing of the international search report
16 SEP 2003

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer
Georgia Helmer
Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 June 2012 (28.06.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/083424 A1

(51) International Patent Classification:
C07K 17/00 (2006.01) *C40B 30/04* (2006.01)
B01D 15/08 (2006.01) *C40B 50/18* (2006.01)
C07K 17/06 (2006.01) *C40B 80/00* (2006.01)
C07K 5/083 (2006.01) *G01N 33/53* (2006.01)
C07K 7/06 (2006.01) *G01N 33/564* (2006.01)
C12N 15/63 (2006.01) *C07K 1/22* (2006.01)

(21) International Application Number:
PCT/CA2011/001366

(22) International Filing Date:
20 December 2011 (20.12.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/425,412 21 December 2010 (21.12.2010) US

(71) Applicant *(for all designated States except US)*: **THE UNIVERSITY OF WESTERN ONTARIO** [CA/CA]; The Gordon Moganson Building, Suite 105, 100 Collip Circle, London, Ontario N6G 4X8 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LI, Shun-Cheng** [CA/CA]; 54 Ambleside Drive, London, Ontario N6G 4M4 (CA). **LI, Xing** [CA/CA]; 35-400 Skyline Avenue, London, Ontario N5X 0B3 (CA). **VOSS, Courtney** [CA/CA]; 227 Grand Avenue, London, Ontario N6C 1M6 (CA).

(74) Agent: **KRUPNIK, Eduardo**; 2010 - 255 Queens Avenue, London, Ontario N6A 5R8 (CA).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: NOVEL AMINO ACID LINKER SEQUENCES FOR LIGAND IMMOBILIZATION

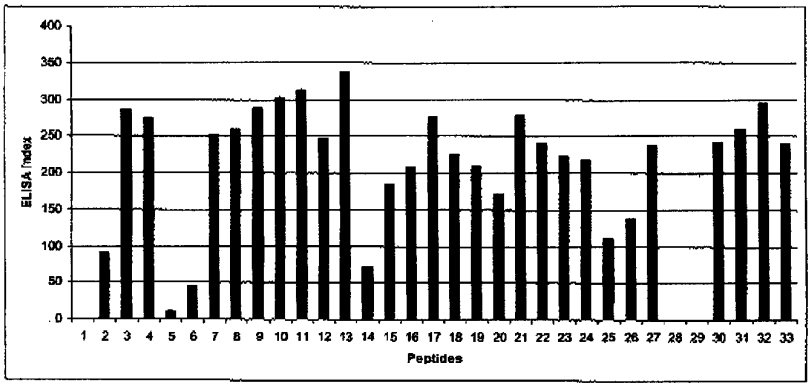


FIGURE 2

(57) **Abstract**: The present invention relates to anchor peptides for linking a ligand molecule to a solid support. The anchor peptide includes an amino acid motif having at least one neutral polar residue and at least one basic polar residue. In one aspect of the invention the anchor peptide is ATKASK (SEQ ID NO: 6). The anchor peptide of the present invention may be used in methods for immobilizing a ligand molecule to a solid support. Ligand molecules include Protein A.

WO 2012/083424 A1

NOVEL AMINO ACID LINKER SEQEUNCES FOR LIGAND IMMOBILIZATION

Field of the Invention

[0001] This invention relates to novel amino acid sequences that can function as linkers or anchors to facilitate immobilization of a protein ligand or another macromolecule onto a solid support.

Background of the Invention

[0002] Protein A is a membrane protein encoded by a bacterium (*Staphylococcus aureus*). Protein A consists of five homologues domains arranged as E-D-A-B-C. Protein A is unique in its ability to bind to most antibodies (including IgG, IgM, IgE, IgA) from various species including humans. Thus, Protein A has been widely used as a protein ligand immobilized on a solid support (eg., resin) to form an affinity column for antibody purification for research and therapeutic applications.

[0003] To fully exploit the antibody-binding property of Protein A for both pharmaceutical and biotechnological applications, methods to improve the immobilization efficiency of Protein A on a solid resin have been under investigation. Immobilization of a protein onto a solid support may be mediated through the ε-amine of the Lys side-chain. Although convenient, this method is non-specific, inefficient and may lead to loss of protein function. US Patent Nos. 5,084,559 and 5,260,373 describe methods of immobilizing Protein A in which a single cysteine residue, or a single arginine residue, is added to the N-terminus of Protein A, or a domain, or multiple domains thereof, to facilitate their immobilization.

[0004] Despite the foregoing, there remains a need to develop improved methods for protein immobilization and enhanced ligand stability.

[0005] This invention focused on identifying short amino acid sequence motifs that may function as site-specific linker or anchor for a protein or another biomolecule. By flanking a Lys residue with polar, neutral amino acids such as Ser and Thr, we have identified a series of novel short peptide sequences that exhibit superior efficiency in coupling the peptide to a solid support.

[0006] These motifs can function as linkers to couple a protein, DNA molecules such as aptamers, or another biomolecule to the surface of a solid support in a site-specific and oriented manner. Immobilized ligands may be used in affinity purification of proteins and other biomolecules for research or therapeutic uses. Depending on the type of molecules immobilized and surface employed, they can also be developed into research or diagnostic tools.

Summary of the Invention

[0007] In one embodiment the present invention provides for an anchor peptide for linking a ligand molecule to a solid support, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

[0008] In another embodiment the present invention provides for a fused product comprising a ligand molecule and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

[0009] In another embodiment the present invention provides for a solid support having an immobilized fused product, the fused product comprising a ligand molecule and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

[0010] In another embodiment the present invention provides for a method of immobilizing a ligand molecule to a solid support comprising: (a) forming a fused product comprising a ligand molecule and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue; and (b) contacting the fused product with a solid support under conditions to immobilize the fused product to the solid support.

[0011] In another embodiment the present invention provides for a method for purifying a molecule of interest in a sample suspected to include the molecule of interest, characterized in that the method comprises: (a) contacting the sample with a fused product immobilized to a solid support, the fused product comprising a ligand molecule

that binds to the molecule of interest and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue, said contacting being under conditions wherein a complex is formed between the molecule of interest and the ligand molecule if said molecule of interest is present in the sample; and (b) recovering the molecule of interest from the complex thereby purifying the molecule of interest from the sample.

[0012] In another embodiment the present invention provides for a method for detecting a molecule of interest in a sample suspected to include the molecule of interest, characterized in that the method comprises: (a) contacting the sample to a fused product immobilized to a solid support, the fused product comprising a ligand molecule that binds to the molecule of interest and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue, said contacting being under conditions wherein a complex is formed between the molecule of interest and the ligand molecule if said molecule of interest is present in the sample; (b) recovering the molecule of interest from the complex; and (c) detecting the recovered molecule of interest.

[0013] In another embodiment the present invention provides for a vector comprising at least one copy of a gene of a ligand molecule and at least one gene capable of coding for an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

[0014] In one embodiment of the present invention the at least one neutral polar residue is serine (S) or threonine (T), and wherein said at least one basic polar residue is lysine (K).

[0015] In one embodiment of the present invention the amino acid motif comprises the formula XX, wherein one X is a K and the other X is S or T.

[0016] In one embodiment of the present invention the anchor peptide further comprises at least one neutral non-polar residue.

[0017] In one embodiment of the present invention the at least one neutral polar residue is S or T, the at least one basic polar residue is K and the at least one neutral non-polar residue is alanine (A).

[0018] In one embodiment of the present invention the amino acid motif comprises at least 2 amino acid residues.

[0019] In one embodiment of the present invention the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 3), wherein each X is K, S or T and wherein U is any amino acid.

[0020] In one embodiment of the present invention the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 4), wherein each X is A, K, S or T and wherein U is any amino acid.

[0021] In one embodiment of the present invention the amino acid motif comprises the formula XX-GA-XX (SEQ ID NO: 5), wherein each X is K, S or T.

[0022] In one embodiment of the present invention the amino acid motif comprises the sequence ATK (SEQ ID NO: 1) or ASK (SEQ ID NO: 2).

[0023] In one embodiment of the present invention the amino acid motif comprises the sequence ATKASK (SEQ ID NO: 6).

[0024] In one embodiment of the present invention the amino acid motif comprises the sequence ATKGATK (SEQ ID NO: 7).

[0025] In one embodiment of the present invention the ligand molecule includes an immunoglobulin-binding protein, an enzyme, a target protein binding partner or a nucleic acid molecule.

[0026] In one embodiment of the present invention the ligand molecule is protein A or a domain of protein A.

[0027] In one embodiment of the present invention the anchor peptide is labelled with a detectable label.

[0028] In one embodiment of the present invention said anchor peptide is labelled with a cytotoxic molecule, a fluorescent molecule or radioactive molecule.

[0029] These and other aspects of the invention will become apparent from the detailed description by reference to the following Figures.

Brief Description of the Figures

[0030] Figure 1 is a table of a peptide library constructed by chemically synthesizing short peptides containing 14-15 amino acid residues with different motifs in accordance to Example 1.

[0031] Figure 2 shows the efficiency of immobilization of different peptides (the sequences of which are provided in Table 1) onto an epoxy-activated 96-well plate as evaluated by an ELISA assay.

[0032] Figure 3 A illustrates the anchor peptide ATKASK (SEQ ID NO: 6) motif linked to a single Z domain. Figure 3 B illustrates eight fusion products (SEQ ID NOS: 11 - 18) according to one embodiment of the present invention. The fusion products labelled 901 and 700 contain the ATKASK (SEQ ID NO: 6) motif.

[0033] Figure 4 illustrates a dot staining analysis of the immobilization efficiency of six of the eight fusion products illustrated in Figure 3 as well as a control bovine serum albumin (BSA) sample on an epoxy resin. On strip A, 901 and 902 fused products were dotted; on strip B, 903 and 904 fused products were dotted; on strip C, 123 fused product was dotted; on strip C, 201 fused product was dotted; and on strip E, control BSA was dotted. Ori, orgininal sample without dilution.

[0034] Figure 5 is a graph showing a BCA protein analysis of the fused products referred to in Figure 4 using a protein assay kit (Pierce Co.). The protein concentrations in those samples were measured. The immobilization efficiency of a protein ligand was

illustrated as percentages of a protein ligand concentration before and after immobilization process.

[0035] Figure 6 panels A to D show the IgG binding capacities of eight affinity columns composed of the protein A ligand (Z) fused to various N-linkers illustrated in Fig 3. A. fusion products 901 and 902, B. fusion products 903 and 904, C. fusion products 285 and 700, and D. fusion products 123 and 201.

Detailed Description of the Invention

[0036] *Definitions*

[0037] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Also, unless indicated otherwise, except within the claims, the use of "or" includes "and" and vice versa. Non-limiting terms are not to be construed as limiting unless expressly stated or the context clearly indicates otherwise (for example "including", "having" and "comprising" typically indicate "including without limitation"). Singular forms including in the claims such as "a", "an" and "the" include the plural reference unless expressly stated otherwise.

[0038] The term "functionally equivalent derivative" is used herein to refer to derivatized versions which retain the function of anchor to immobilize the selected protein without substantial loss of protein function.

[0039] The term "ligand" as used herein refers to a molecule or group of molecules capable of binding to one or more specific sites of a receptor, another molecule or target analyte. There can exist more than one ligand for a given target. The ligands may differ from one another in their binding affinities for the target molecule.

[0040] *Overview of the Invention*

[0041] The present invention relates in general to an anchor peptide. The anchor peptide may be used for linking a ligand molecule to a solid support. The anchor peptide

may comprise an amino acid motif having at least one neutral polar residue and at least one basic polar residue. In one embodiment of the present invention the anchor peptide may further comprise at least one neutral non-polar residue. The neutral polar residue may be a threonine (T) or a serine (S) residue, the basic polar residue may be a lysine (K) residue, and the neutral non-polar residue may be an alanine (A) residue.

[0042] In another aspect, the present invention provides for an anchor peptide for linking a ligand molecule to a solid support, said anchor peptide comprising an amino acid motif having at least one neutral polar residue, at least one basic polar residue and at least one neutral non-polar residue.

[0043] The present method is not particularly restricted with respect to the ligand molecule that may be immobilized, and is therefore useful for immobilizing any protein of interest. Of particular interest is use of the method to immobilize immunoglobulin-binding proteins, including but not limited to, Protein A, Protein G, Protein A/G and Protein L. Other ligand molecules of interest may include nucleic acid ligands, including aptamers.

[0044] *The Anchor Peptide*

[0045] The anchor peptide may be suitable to immobilize the selected ligand molecule onto a solid support while substantially retaining the function of the selected ligand molecule, i.e. retaining at least about 50% of ligand molecule function, such as immunoglobulin binding, in its immobilized state. In one embodiment of the present invention, the anchor or linker may comprise a motif having at least 2 amino acid residues. The amino acid motif of the anchor may comprise equally three groups of amino acids- neutral non-polar, neutral polar, basic polar residues. An anchor sequence may also contain only the neutral-polar and basic residues in different combinations and arrangements.

[0046] In one embodiment of the present invention the anchor peptide may comprise an amino acid sequence ATK (SEQ ID NO. 1) or ASK (SEQ ID NO. 2).

[0047] In one embodiment the present invention provides for a novel peptide including an amino acid formula XX-UU-XX (SEQ ID NO. 3), wherein each X may be K, S or T and U may be any amino acid. In one embodiment of the present invention the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 4), wherein each X is A, K, S or T and wherein U is any amino acid. In another embodiment the formula may be XX-GA-XX (SEQ ID NO. 5) wherein each X may be K, S or T.

[0048] In one embodiment of the invention, the anchor peptide may comprise an amino acid sequence ATKASK (SEQ ID NO. 6), or a functionally equivalent derivative thereof including at least 50% sequence homology.

[0049] In one embodiment the present invention provides for a novel peptide including the amino acid sequence SEQ ID NO. 6.

[0050] In one embodiment of the invention, the anchor peptide may comprise an amino acid sequence ATKGATK (SEQ ID NO. 7), or a functionally equivalent derivative thereof including at least 50% sequence homology.

[0051] In one embodiment the present invention provides for a novel peptide including the amino acid sequence SEQ ID NO. 7.

[0052] The anchor peptides of the present invention may also be provided as a fused product comprising the anchor peptide and a ligand molecule. The anchor peptide may be an N- or C-terminal extension of a ligand protein.

[0053] The anchor peptide or the fused product comprising the anchor peptide and a ligand protein may be synthesized by any known method in the art of peptide synthesis including solid phase synthesis (Merrifield (1964) J. Am. Chem. Assoc. 65:2149; J. Amer. Chem. Soc. 85:2149 (1963); and Int. J. Peptide Protein Res. 35:161-214 (1990)) or synthesis in homogenous solution (Methods of Organic Chemistry, E. Wansch (Ed.) Vol. 15, pts. I and II, Thieme, Stuttgart (1987) to generate synthetic peptides.

[0054] The anchor peptide or the fused product comprising the anchor peptide linked to a ligand protein may be made by recombinant methods known in the art. In this regard, the anchor peptide may generally be bound or fused to a terminus of the selected ligand protein such that the function of the protein may substantially be retained. Thus, depending on the ligand protein, the anchor may be bound to the N-terminus or the C-terminus of the selected ligand protein. In certain aspects, the anchor peptide may be bound to both termini, or may additionally be bound to an amino acid residue which may not be a terminal residue. In another aspect, the anchor peptide may be bound to a solid support prior to the coupling of the protein ligand by chemical or recombinant means or through the use of an enzyme.

[0055] The anchor peptides of the invention may be labelled with a label to facilitate their detection in an assay as is understood by one of skill in the art. Such labels may include but are not limited to radioactive label, fluourescent label or cytotoxic label. The anchor peptides of the invention may be provided with a carrier such as for example coupled to bovine serum albumin (BSA) or keyhole limpet haemocyanin. The anchor peptides may be preceded by a Biotin-GGYG (SEQ ID NO: 55) N-terminal sequence that may facilitate peptide concentration determination by OD280 (of Tyr or Y) measurement. The N-terminal biotin moiety may provide a ligand for ELISA assay (see Figure 2) using avidin-HRP.

[0056] Once the ligand molecule is bound or fused to the anchor peptide of the present invention, the fused product may be immobilized onto a solid support. Suitable solid supports include epoxy-based supports, e.g. beads or resins, or solid supports having the surface chemistry of an epoxy group, or materials that are functionalized on the surface to allow coupling of a protein via the anchor sequence.

[0057] To immobilize the fused product to the solid support, the solid support may be treated in an aqueous solution. The aqueous solution may include a binding buffer, such as 0.1 M sodium phosphate/0.5 M Na2SO4, pH, e.g. in the range of about 6.5 to about 7.2. In one aspect the pH is 6.8. Suitable salts for use in this regard may include salts that may be compatible with the solid support and fused product, non-toxic

with regard to the intended use of the immobilized protein and otherwise suitable for use. The solid support may be mixed with the fused product, for example at room temperature, for a sufficient period of time for immobilization to occur.

[0058] In embodiment, the present invention relates to a matrix for affinity separation. One end of the anchor peptides of the present invention may be bound to the matrix, and the other end of the anchor peptide may be bound to the ligand molecule. The matrix of the present invention, when compared to a matrix having the ligand directly bound to the matrix, may exhibit an increased immobilization capacity during two or more separations.

[0059] Examples of the material for the solid support is based on polymers having hydrophilic surface, such as polymers having hydroxyl group (-OH), carboxyl group (-COOH), aminocarbonyl group (-$CONH_2$, possibly in N-substituted forms), amino group (-NH_2, possibly in substituted form), oligo or polyethyleneoxy group on their external surface and if present, also on internal surface. In one embodiment, the polymer may be a synthetic polymer (e.g., polymethacrylate, polyacrylamide, or styrene-divinylbenzene copolymer). Such a synthetic polymer may be easily produced by a known method (see J. MATER. CHEM 1991, 1(3), 371-374, for example). Alternatively, a commercially available product such as TOYOPEARL™ may also be used. In another embodiment, the polymer is a polysaccharide (e.g., dextran, starch, cellulose, pullulan, or agarose). Such a polysaccharide may be easily produced by a known method (see Japanese Patent No. 4081143, for example). Alternatively, a commercially available product such as Sepharose (manufactured by GE Healthcare Biosciences) may also be used. In still another embodiment, an inorganic support (e.g., silica or zirconium oxide) may be used.

[0060] The solid support may be in the shape of particles. The particles may be porous or nonporous. The solid support in the shape of particles may be used as a packed bed, or may be used in a suspended form. The suspended form may be an expanded bed or a pure suspension, in which the particles can move freely. When using a packed bed, or an expanded bed, separation process used in known affinity chromatographic methods is used. When using a pure suspension, a batch method is used.

[0061] The solid support in the shape of particles according to this embodiment may preferably have a particle size (volume average particle size) of about 20 to about 200 micrometers, and more preferably about 30 to about 100 micrometers. If the solid support has a particle size of less than about 20 micrometers, the pressure of column packed with this support may increase to an impractical level at a high flow rate. If the solid support has a particle size of more than about 200 micrometers, the binding capacity may deteriorate. Note that the term "particle size" used herein refers to a volume average particle size determined using a laser diffraction/scattering particle size distribution analyzer.

[0062] The solid support according to this embodiment may preferably be porous, and having a specific surface area larger than about 50-1000 m^2/g-, and more preferably about 80-400 m^2/g. If the affinity chromatography packing material has a specific surface area of less than about 50 m^2/g, the binding capacity may deteriorate. If the affinity chromatography packing material has a specific surface area of more than about 1000 m^2/g, the packing material may be destroyed at a high flow rate due to a decrease in strength, so that the column pressure may increase. Note that the term "specific surface area" used herein refers to a value of the surface area of pores (pore size: 10 to 5000 nm) determined using a mercury porosimeter by the dry weight of the particles.

[0063] The solid support according to this embodiment may preferably have a volume average pore size of about 50 to about 400 nm, and more preferably about 100 to about 300 nm. If the affinity solid support has a volume average pore size of less than about 50 nm, the binding capacity may significantly deteriorate at a high flow rate. If the affinity chromatography packing material has a volume average pore size of more than about 400 nm, the binding capacity may deteriorate irrespective of the flow rate. Note that the term "volume average pore size" used herein refers to the volume average pore size of pores (pore size: 10 to 5000 nm) determined using a mercury porosimeter.

[0064] Specific examples of porous particles used as the solid support according to this embodiment may include porous organic polymer particles including a copolymer of 20 to 50 wt% of a crosslinkable vinyl monomer, 3 to 80 wt% of an epoxy group-

containing vinyl monomer, and 20 to 80 wt% of a diol group-containing vinyl monomer, and having a particle size of 20 to 80 micrometers, a specific surface area of 50 to 150 m^2/g, and a volume average pore size of 100 to 400 nm.

[0065] In yet another form embodiment, the solid support may be in another form such as a surface, a monolith, a chip, capillaries, or a filter.

[0066] The ligand molecule may be attached to the support via conventional coupling techniques utilizing, e.g. amino and/or carboxy groups present in the ligand. Bisepoxides, epichlorohydrin, CNBr, N-hydroxysuccinimide (NHS) etc. are well-known coupling reagents. Between the support and the ligand, a molecule known as a spacer may be introduced, which will improve the availability of the ligand and facilitate the chemical coupling of the ligand to the support. Alternatively, the ligand may be attached to the support by non-covalent bonding, such as physical adsorption or biospecific adsorption.

[0067] In one embodiment of the present invention, the molecule ligands may be peptide ligands that may be used in the purification of immunoglobulins of interest from a composition comprising at least one immunoglobulin of interest and at least one other substance from which the immunoglobulin is to be separated.

[0068] The fused product comprising the anchor peptides of the present invention bound to peptide ligands, may be immobilized on to a solid support or may be used in solution. The solid support may then be contacted with a composition containing an immunoglobulin, the immunoglobulin then binding to the peptide ligand.

[0069] When using the fused products immobilized on a solid support, the composition containing the immunoglobulin may be contacted to the fused products-solid support either before or after the solid support material is packed into a column. For example, the fused products-solid support material may be incubated with the composition containing the immunoglobulin and then this mixture may be packed into a column for further purification steps. Alternatively, the fused product-solid support may be first packed into the column followed by the addition of the composition containing

the immunoglobulin of interest. The fused products-solid support material may first be packed in the column followed by the introduction of the composition containing the immunoglobulin of interest.

[0070] The recovery of the immunoglobulin from the fused products-solid support may involve any technique conventionally used in affinity chromatography to remove compounds from solid supports, including changes in pH or ionic strength. In one preferred embodiment of this invention, the immunoglobulin is eluted using a buffer at or below pH 4.

[0071] In an additional embodiment of the invention, the fused products may be used in assays for the detection of molecules of interest such as antibodies in a solution or composition suspected of containing the same. The fused products may be coupled to a detectable group or label. The contacting and separating steps may be carried out in solution or with immobilized fused products. Methods of detecting labelled fused products include those for the detection of labelled immunoglobulin. When carried out in solution, fused products bound immunoglobulin may be separated and/or detected from the solution using a variety of techniques including, but not limited to, fluorescence-activated cell sorting, membrane filtration and mass spectrometry.

[0072] One embodiment of this invention includes methods of using a fused product of the present invention in methods for determining the presence or concentration of a molecule of interest in a sample. The ligand component of the fused product being a lingand of the molecule of interest. The method may include:

[0073] (a) contacting said sample to the fused product under conditions wherein a molecule of interest/ligand complex is formed if said molecule of interest is present in the sample; and

[0074] (b) detecting of the molecule of interest by detecting the molecule of interest/ligand complex, thereby determining the presence of the molecule of interest in the sample. Step (b) may also include measuring the amount of the molecule of interest in the sample by quantitatively detecting the complex.

[0075] One embodiment of this invention includes methods of using a fused product of the present invention in an affinity column for the determination of presence and/or concentration of molecules of interest in a sample suspected of containing the same. The methods may include the following steps:

[0076] (a) immobilizing a fused product of the present invention to an affinity column, the ligand component of the fused product being a ligand of the molecule of interest;

[0077] (b) running the sample through the affinity column under conditions whereby the molecule of interest is captured by the fused product if said molecule of interest is present in the sample, thereby ;

[0078] (c) recovering the molecule of interest from the column with a recovering agent; and

[0079] (d) measuring the quantity of molecule of interest captured by the column by methods such as direct fluorescence measurement, high performance liquid chromatography and mass spectrometry of the molecule of interest. Alternatively, prior to the recovering step, the molecule of interest/ligand complex formation within the column may be detected by the use of fluorescence, or fluorescence in combination with quenchers, or fluorescence polarization, and electro-affinity analysis.

[0080] Embodiments of the invention are described in the following specific example which is not to be construed as limiting.

EXAMPLES

Example 1 – Chemical synthesis of a peptide library

[0081] A peptide library was constructed by chemically synthesizing short peptides containing 14-15 amino acid residues with different motifs (Figure 1). These peptides were synthesized by the Fmoc chemistry based solid-phase peptide synthesis and labelled at the N-terminus with biotin. After synthesis, the peptides were dissolved in

water and the identities (eg., molecular masses) were confirmed by MALDI-TOF mass analysis.

Peptides 1 to 24 of Figure 1 are designed in such a way that a lysine (K) is next to an Ser (S) or Thr (T) residue or flanked by different combinations of serine (S), threonine (T) and other types of amino acids. The motif length may be six as in XX-UU-XX (SEQ ID NO: 3) or XX-GA-XX (SEQ ID NO: 5) (peptides 9 to 16, wherein at least one K is in the X positions). The GA dipeptide may serve as a spacer between the XX motif, but GA can be replaced by other amino acid types as in the motif XX-UU-XX (SEQ ID NO: 3), where U represents any amino acid. The motif can be as short as containing two amino acid as in the motif XX, wherein one X is K, and the other X is S or T. The peptide core sequence may be preceded by a Biotin-GGYG (SEQ ID NO: 55) N-terminal sequence that may facilitate peptide concentration determination by OD280 (of Tyr or Y) measurement. The N-terminal biotin moiety may provide a ligand for ELISA assay (see Figure 2) using avidin-HRP.

Example 2 – Identification of peptide motifs that allow efficient ligand immobilization onto a solid support

[0082] The synthetic peptides are each dissolved in ddH_2O and the pH value of the solution adjusted to pH 7-8 using a 0.1N NaOH solution. The resulting peptide solutions are added, respectively, to different wells of a 96-well plate containing epoxy groups on the surface and incubated at 4 °C overnight. The plate is washed in PBS and then blocked with 5% skim milk at room temperature for 1 hour. After blocking, a avidin-HRP solution (1:40,000 dilution) was added to the plate and incubated at room temperature for 1 hour. The plate is washed again three times with PBS and incubated with tetramethylbenzidine (TMB), a colorless substrate. After 20 minutes, the reaction is stopped by adding 2M H2SO4. The OD at 405 nm was recorded for the plate using a Multi-label reader (Perkin-Elmmer). As shown in Figure 2, a higher OD405 value corresponds to a greater amount of a peptide immobilized in a well. As illustrated in Figure 2, a number of peptides showed efficient ligand immobilization to the solid support. Peptides that exhibited the greatest immobilization efficiency contained the ATK (SEQ ID NO: 1) and ASK (SEQ ID NO: 2) motifs (highlighted in Figure 1).

Example 3 – Protein ligand immobilization on the epoxy resin

[0083] Using a molecular cloning method, a single Z domain was fused with various N-linker sequences as illustrated in Fig. 3. These fused protein ligands were expressed in a bacterial expression system. The target proteins were purified with the Ni-NTA affinity column (GE Healthcare) to highly homology. Protein ligand's purity was examined by SDS-PAGE analysis.

[0084] The purified protein ligands were prepared in the immobilization buffer (0.1 M sodium phosphate/0.5 M Na2SO4, pH 6.8). Five milligram of each protein ligand solution was mixed with 1 millilitre of the epoxy-activated resin (JSR). The protein ligand/resin mixture was incubated at about room temperature overnight with a rotating mixer. Immobilization of the fused product in each case may be monitored by the dot staining method (FIG. 4) and by the BCA protein measuring kit (FIG. 5).

[0085] The results illustrated in FIGS. 4 and 5 confirmed that the Z protein ligand fused with a N-linker containing either "ATKASK" (SEQ ID NO: 6) (ligand 901) motif or "ATKGATK" (SEQ ID NO: 7) (ligand 902) motif was successfully immobilized on the beads, while the Z ligand fused with other N-linker sequences without the motifs "ATKASK" (SEQ ID NO: 6) or "ATKGATK" (SEQ ID NO: 7) were less efficient to be immobilized onto the beads under the same conditions.

Example 4 – IgG binding capacity of Immobilized Protein A containing an ATKASK (SEQ ID NO: 6) or ATKGATK (SEQ ID NO: 7) anchor

[0086] Eight affinity columns of 1 ml bed volume, packed with the epoxy beads (JSR) immobilized with eight protein ligands with various N-linkers, were analyzed to determine their human IgG binding capacity (Fig. 5). The IgG solution at 1 mg/ml of PBS was added to each affinity column. The flow through samples were collected. OD280 values were monitored between the loading IgG solution and the flow through samples. The differences of the OD280 values between the initial IgG solution and the flow through samples are the indication of the affinity column's IgG binding capacity. The higher an affinity column's IgG binding capacity, the lower the percentage of the

initial IgG OD280 reading and the flow through sample OD280 reading. Results are depicted in Figure 6. Figure 6 A and 6 C show that in both 901 and 902 columns as well as 700 columns, the percentage of eluted IgG (unbound IgG) was approximately at 25-30 % at the 20 mg IgG loading, while the rest of the columns showed the percentage above 50%. These three columns contain the protein ligands fused with the N-linker containing the motif of "ATKASK" (SEQ ID NO: 6) (901 and 700) or the motif of "ATKGATK" (SEQ ID NO: 7) (902). These data indicate that the N-linker containing the motif of "ATKASK" (SEQ ID NO: 6) would greatly facilitate a protein ligand to be immobilized on an epoxy bead and thus enhance the IgG binding capacity.

Methods and Materials for Examples 3 and 4

[0087] **(1) Vector constructs for Protein A expression** – The Z domain of protein A fused to a different anchor peptide was cloned into the pETM11 vector (European Molecular Biology Laboratory) or pET302/N-His (Invitrogen). The original Protein A gene was amplified by PCR method using a TAP-Tag plasmid (EMBL) containing Protein A gene as a template. The Z domain was created by the PCR assisted QuickChange Site-Directed Mutagenesis method (Stratagene Cloning System).

[0088] All protein A ligands with various N-linkers listed in FIG. 3B (201, 123, 901, 902, 903, 904, 285, 700) were manufactured in the E. coli expression system, BL21(DE3) pLysS competent cells, as described in the following paragraphs.

[0089] The eight expression vectors listed in Fig. 3B were constructed as follows:

[0090] 201 expression vector was constructed by PCR the single Z domain with the restriction enzyme sites: EcoRI/BamHI and cloned the single Z domain into the expression vector of pET302/N-His (Invitrogen).

[0091] 123 expression vector was constructed by PCR the single Z domain with the restriction enzyme sites: NcoI/BamHI and cloned the single Z domain into the expression vector of pETM11 (EMBL).

[0092] 901, 902, 903, and 904 expression vectors were constructed by deleting V and introducing ATKASK (SEQ ID NO: 6) (901), ATKGATK (SEQ ID NO: 7) (902), AHKGAHK (SEQ ID NO: 56) (903) and GHTKGHTK (SEQ ID NO: 57)(904) into the 123 by using the QuickChange Site-Directed Mutagenesis method (Stratagene Cloning system).

[0093] 285 expression vector was constructed by deleting MSDYDIPTT (SEQ ID NO: 58) in 123 by using the QuickChange Site-Directed Mutagenesis method (Stratagene Cloning system).

[0094] 700 expression vector was constructed by deleting MV and adding ATKASK (SEQ ID NO: 6) in 285 by using the QuickChange Site-Directed Mutagenesis method (Stratagene Cloning system).

[0095] **(2) Recombinant protein expression and purification** - BL21(DE3) pLysS competent cells allow high-efficiency protein expression of any gene that is under the control of a T7 promoter and has a ribosome binding site. BL21(DE3)pLysS is lysogenic for λ-DE3, which contains the T7 bacteriophage gene I, encoding T7 RNA polymerase under the control of the lac UV5 promoter. BL21(DE3)pLysS also contains a plasmid, pLysS, which carries the gene encoding T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG. Target DNA plasmids were used to transfect BL21 competent cells. 0.5 µl of plasmid DNA containing about 0.5 ug of DNA was used for transfection. The BL21 competent cells were stored at – 80 degree Celsius. An aliquot of BL21 cells (20 µl) was thawed and put on ice. 0.5 µl of plasmid DNA was mixed with the competent cells and put on ice for 60 minutes. After incubation on ice, these cells containing plasmid DNA were "heat shocked" in 45 degree C water batch for 45 seconds and immediately put on ice for another 2 minutes. Then, 30 µl of LB medium was added to the bacterial cells and shaken at 37 degree C for 1 hour. These cells were then spread on agar /LB plates containing kanamycine (final concentration: 100 µg/ml). These plates were incubated at 37 degree C overnight.

[0096] For expression of recombinant Protein A domains, the DNA plasmids were transformed into E. coli BL21. Overnight cultures were inoculated into fresh medium and grown to an OD600nm of approximately 0.5. Isopropyl beta-D-1-thiogalactopyranoside was added to a concentration of 1 mM, and the culture was grown for a further 15 hours at 18 degree C. Cells were harvested by centrifugation at 7,000 rpm for 10 min in a Sorvall GS-3 rotor.

[0097] The pellet was resuspended in phosphate-buffered saline (PBS) containing protease inhibitor (Roche Applied Science), lysozyme (200 µg/ml), and DNase I (3 µg/ml). Cells were lysed by repeated freeze/thaw cycles. Cell debris was removed by centrifugation.

[0098] Recombinant proteins expressed in pETM11 vector contained an amino-terminal His-Tag. His-Tag fusion proteins were purified using the Ni-NTA columns (Qiagen) and dialyzed against PBS.

[0099] **(3) Protein Analysis by Polyacrylamide gel electrophoresis (SDS-PAGE)** - Electrophoresis was performed on polyacrylamide gels in the presence of sodium dodecyl sulphate (SDS). Samples (10 µl) were mixed with 10 µl of sample buffer (62.5mM Tris-HCl, pH 6.8, 2% SDS, 5% mercaptoethanol, 10% glycerol and 0.0025% brophenol blue), heated at 90 C for 2 min, cooled and loaded on gels. Electrophoresis was performed at 75 mA in an apparatus purchased from Bio-Rad. Gels were stained in a solution of 0.5 mg/ml Coomassie blue in 5:5:1 methanol/water/acidic acid and de-stained in water heated in a microwave.

[00100] **(4) Peptide synthesis** -Fmoc solid-phase peptide synthesis method was used to synthesize all peptides. Protected amino acid derivatives were purchased from AnaSpec (US); N,N-diisopropylethylamine (DIEA), HBTU, HOBT, piperidine were purchased from EMD (Canada); and Tental gel resin was purchased from Intavis (Canada). Other reagents and solvents were purchased from Sigma (Canada). MALDI-TOF-mass spectra were recorded on a Micromass® bench-top MALDI-TOF mass analysis with reflection ion mode.

[00101] **(5). Synthesis of porous beads** - 8.2 g of glycidyl methacrylate (manufactured by Mitsubishi Rayon Co., Ltd.), 65.9 g of trimethylolpropane trimethacrylate (manufactured by Sartomer), and 90.6 g of glycerol monomethacrylate (manufactured by NOF Corporation) were dissolved in 245.8 g of 2-octanone (manufactured by Toyo Gosei Co., Ltd.) and 62 g of acetophenone (manufactured by Wako Pure Chemical Industries, Ltd.), and 2 g of 2,2'-azoisobutyronitrile (manufactured by Wako Pure Chemical Industries, Ltd.) was added to the solution to prepare an organic monomer solution.

[00102] 8.5 g of polyvinyl alcohol ("PVA-217" manufactured by Kuraray Co., Ltd.), 0.43 g of sodium dodecyl sulfate ("Emal 10G" manufactured by Kao Corporation), and 21.3 g of sodium sulfate (manufactured by Wako Pure Chemical Industries, Ltd.) were added to 4240 g of distilled water. The mixture was stirred overnight to prepare an aqueous solution.

[00103] A separable flask (7 l) was charged with the resulting aqueous solution. The separable flask was equipped with a thermometer, a stirring blade, and a cooling tube, and placed on a hot water bath. The aqueous solution was stirred at 600 rpm in a nitrogen atmosphere. When the temperature of the aqueous solution had reached 85°C, the organic monomer solution was added to the aqueous solution using a dropping funnel. The mixture was stirred for 5 hours.

[00104] After cooling the reaction solution, the reaction solution was transferred to a polypropylene bottle (5 l). The reaction solution was allowed to stand until the beads floated. Unnecessary water was then sucked out from the bottom of the bottle. Acetone was then added to the reaction solution to precipitate the beads. After allowing the reaction solution to stand for 3 minutes, acetone was removed by decantation. After repeating this operation twice, water was added to precipitate the beads. After allowing the mixture to stand for 3 minutes, the mixture was subjected to decantation. After replacing the dispersion medium of the bead dispersion with acetone, the mixture was air-dried overnight. The mixture was then dried using a vacuum dryer to obtain 90 g of

porous beads (hereinafter referred to as "PB"). The average particle size of the PB was 43 micrometers, and the specific surface area of the PB was 83 m^2/g.

[00105] **(6) Immobilization of protein ligands on the epoxy beads-** Epoxy-activated beads obtained as described in the previous section, were used for immobilization. The epoxy beads were washed with the immobilization buffer before the immobilization. Five milligram of protein ligand fused to an anchor peptide was added to 1 ml of the epoxy resin. The protein/resin mixture was incubated at room temperature overnight with a rotating shaker. After overnight immobilization, the protein ligand fused to an anchor peptide (such as 901, 902, etc) resin was blocked with 1M thioglyceol in 0.1 M Na2SO4, pH 8.3 for 4 hours at room temperature. After blocking, the resin was washed with 1 x PBS (+ 0.5% Tween 20). After washing, a protein A affinity resin was ready for further use.

[00106] **(7) Dot analysis** – Referring to Figure 4, products 901 (Figure 4 A), 902 (Figure 4 A), 903 (Figure 4 B), 904 (Figure 4 B), 123 (Figure 4 C) and 201 (Figure 4 D) before immobilization (top row) and after immobilization (product left in solution, bottom row) on the epoxy beads were diluted 1:1 in PBS respectively and dotted on a nitrocellulose filter paper. The filter strips were stained by Coomassie blue. In a control experiment, BSA (bovine serum albumin, Pierce Co.) was immobilized on the same epoxy beads and samples were taken before and after the immobilization reaction and dotted on a nitrocellulose filter paper, as illustrated in Figure 4 E (Top row: before immobilization; bottom row: after immobilization).

(8) BCA protein analysis – Referring to Figure 5 a BCA protein analysis of the fused products referred to in Figure 4 was performed using a protein assay kit (Pierce). Briefly, BCA Protein Assay combines the well-known reduction of Cu2+ to Cu1+ by protein in an alkaline medium with the highly sensitive and selective colorimetric detection of the cuprous cation (Cu1+) by bicinchoninic acid. The first step is the chelation of copper with protein in an alkaline environment to form a light blue complex. In this reaction, known as the biuret reaction, peptides containing three or more amino acid residues form a colored chelate complex with cupric ions in an alkaline environment containing sodium

potassium tartrate. In the second step of the color development reaction, bicinchoninic acid (BCA) reacts with the reduced (cuprous) cation that was formed in step one. The intense purple-colored reaction product results from the chelation of two molecules of BCA with one cuprous ion. The BCA/copper complex is water-soluble and exhibits a strong linear absorbance at 562 nm with increasing protein concentrations.

[00107] For the BCA assay, 200 µl of the reaction solution was added into each well on a 96-well plate. Then, 25 µl of BSA standard samples or unknown protein samples was added into each well. The 96-well plate containing samples was incubated at 37 °C, for 30 minutes. The purple color was measured at OD595nm using the microplate reader (Perkin-Elmer). A BSA standard curve was constructed by using protein concentration as X axis and OD595 as Y axis. Unknown protein samples' concentrations were deduced based on the BSA standard curve.

[00108] The protein concentrations in those samples were measured. The immobilization efficiency of a protein ligand was illustrated as percentages of a protein ligand concentration before and after immobilization process.

[00109] **(9) IgG binding capacity of a Protein A affinity column**-Protein A affinity columns were evaluated for their IgG binding capacities. Human IgG solution at 1 mg/ml of PBS was loaded on 1 ml of bed volume Protein A affinity resin. The flow-through samples were collected. OD280 value was measured in both loading IgG solution and flow-through samples to calculate the IgG content in samples. The percentage of IgG content in the loading solution and in the flow-through samples indicated a Protein-A affinity column's IgG binding capacity (illustrated at Y axis of Figure 6).

CLAIMS

We Claim:

1. An anchor peptide for linking a ligand molecule to a solid support, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

2. The anchor peptide of claim 1, characterized in that said at least one neutral polar residue is serine (S) or threonine (T), and wherein said at least one basic polar residue is lysine (K).

3. The anchor peptide of claim 1, characterized in that the amino acid motif comprises the formula XX, wherein one X is a K and the other X is S or T.

4. The anchor peptide of claim 1, characterized in that said anchor peptide further comprises at least one neutral non-polar residue.

5. The anchor peptide of claim 4, characterized in that said at least one neutral polar residue is S or T, said at least one basic polar residue is K and said at least one neutral non-polar residue is alanine (A).

6. The anchor peptide of claim 1, characterized in that the amino acid motif comprises at least 2 amino acid residues.

7. The anchor peptide of claim 1, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 3), wherein each X is K, S or T and wherein U is any amino acid.

8. The anchor peptide of claim 1, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 4), wherein each X is A, K, S or T and wherein U is any amino acid.

9. The anchor peptide of claim 1, characterized in that the amino acid motif comprises the formula XX-GA-XX (SEQ ID NO: 5), wherein each X is K, S or T.

10. The anchor peptide of claim 1, characterized in that the amino acid motif comprises the sequence ATK (SEQ ID NO: 1) or ASK (SEQ ID NO: 2).

11. The anchor peptide of claim 1, characterized in that the amino acid motif comprises the sequence ATKASK (SEQ ID NO: 6).

12. The anchor peptide of claim 1, characterized in that the amino acid motif comprises the sequence ATKGATK (SEQ ID NO: 7).

13. The anchor peptide of claim 1, characterized in that the ligand molecule includes an immunoglobulin-binding protein, an enzyme, a target protein binding partner or a nucleic acid molecule.

14. The anchor peptide of claim 1, characterized in that the ligand molecule is protein A or a domain of protein A.

15. The anchor peptide of claim 1, characterized in that the anchor peptide is labelled with a detectable label.

16. The anchor peptide of claim 1, characterized in that said anchor peptide is labelled with a cytotoxic molecule, a fluorescent molecule or radioactive molecule.

17. A fused product comprising a ligand molecule and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

18. The fused product of claim 17, characterized in that said at least one neutral polar residue is serine (S) or threonine (T), and wherein said at least one basic polar residue is lysine (K).

19. The fused product of claim 17, characterized in that the amino acid motif comprises the formula XX, wherein one X is a K and the other X is S or T.

20. The fused product of claim 17, characterized in that said anchor peptide further comprises at least one neutral non-polar residue.

21. The fused product of claim 20, characterized in that said at least one neutral polar residue is S or T, said at least one basic polar residue is K and said at least one neutral non-polar residue is alanine (A).

22. The fused product of claim 17, characterized in that the amino acid motif comprises at least 2 amino acid residues.

23. The fused product of claim 17, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 3), wherein each X is K, S or T and wherein U is any amino acid.

24. The fused product of claim 17, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 4), wherein each X is A, K, S or T and wherein U is any amino acid.

25. The fused product of claim 17, characterized in that the amino acid motif comprises the formula XX-GA-XX (SEQ ID NO: 5), wherein each X is K, S or T.

26. The fused product of claim 17, characterized in that the amino acid motif comprises the sequence ATK (SEQ ID NO: 1) or ASK (SEQ ID NO: 2).

27. The fused product of claim 17, characterized in that the amino acid motif comprises the sequence ATKASK (SEQ ID NO: 6).

28. The fused product of claim 17, characterized in that the amino acid motif comprises the sequence ATKGATK (SEQ ID NO: 7).

29. The fused product of claim 17, characterized in that the ligand molecule includes an immunoglobulin-binding protein, an enzyme, a target protein binding partner or a nucleic acid molecule.

30. The fused product of claim 17, characterized in that the ligand molecule is protein A or a domain of protein A.

31. The fused product of claim 17, characterized in that the fused product is labelled with a detectable label.

32. The fused product of claim 17, characterized in that the fused product is labelled with a cytotoxic molecule, a fluorescent molecule or radioactive molecule.

33. The fused product of claim 17, characterized in that the ligand molecule is a protein and said anchor peptide is bound to the N-terminus or the C-terminus of the ligand protein.

34. A solid support having an immobilized fused product, the fused product comprising a ligand molecule and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

35. The solid support of claim 34, characterized in that said at least one neutral polar residue is serine (S) or threonine (T), and wherein said at least one basic polar residue is lysine (K).

36. The solid support of claim 34, characterized in that the amino acid motif comprises the formula XX, wherein one X is a K and the other X is S or T.

37. The solid support of claim 34, characterized in that said anchor peptide further comprises at least one neutral non-polar residue.

38. The solid support of claim 37, characterized in that said at least one neutral polar residue is S or T, said at least one basic polar residue is K and said at least one neutral non-polar residue is alanine (A).

39. The solid support of claim 34, characterized in that the amino acid motif comprises at least 2 amino acid residues.

40. The solid support of claim 34, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 3), wherein each X is K, S or T and wherein U is any amino acid.

41. The solid support of claim 34, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 4), wherein each X is A, K, S or T and wherein U is any amino acid.

42. The solid support of claim 34, characterized in that the amino acid motif comprises the formula XX-GA-XX (SEQ ID NO: 5), wherein each X is K, S or T.

43. The solid support of claim 34, characterized in that the amino acid motif comprises the sequence ATK (SEQ ID NO: 1) or ASK (SEQ ID NO: 2).

44. The solid support of claim 34, characterized in that the amino acid motif comprises the sequence ATKASK (SEQ ID NO: 6).

45. The solid support of claim 34, characterized in that the amino acid motif comprises the sequence ATKGATK (SEQ ID NO: 7).

46. The solid support of claim 34, characterized in that the ligand molecule includes an immunoglobulin-binding protein, an enzyme, a target protein binding partner or a nucleic acid molecule.

47. The solid support of claim 34, characterized in that the ligand molecule is protein A or a domain of protein A.

48. The solid support of claim 34, characterized in that the fused product is labelled with a detectable label.

49. The solid support of claim 34, characterized in that fused product is labelled with a cytotoxic molecule, a fluorescent molecule or radioactive molecule.

50. The solid support of claim 34, characterized in that the ligand molecule is a protein and said anchor peptide is fused to the N-terminus or the C-terminus of the ligand protein.

51. The solid support of claim 34, characterized in that the solid support includes porous or nonporous particles.

52. The solid support of claim 51, characterized in that the particles are packed in a column or suspended in a solution.

53. The solid support of claim 34, characterized in that the solid support is in the form of a surface, a monolith, a chip, a capillary or a filter.

54. A method of immobilizing a ligand molecule to a solid support comprising:

a. forming a fused product comprising a ligand molecule and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue; and
b. contacting the fused product with a solid support under conditions to immobilize the fused product to the solid support.

55. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that said at least one neutral polar residue is serine (S) or threonine (T), and wherein said at least one basic polar residue is lysine (K).

56. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that said anchor peptide further comprises at least one neutral non-polar residue.

57. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that said at least one neutral polar residue is S or T, said at least one basic polar residue is K and said at least one neutral non-polar residue is alanine (A).

58. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the amino acid motif comprises at least 2 amino acid residues.

59. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 3), wherein each X is K, S or T and wherein U is any amino acid.

60. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the amino acid motif comprises the formula XX-UU-XX (SEQ ID NO: 4), wherein each X is A, K, S or T and wherein U is any amino acid.

61. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the amino acid motif comprises the formula XX-GA-XX (SEQ ID NO: 5), wherein each X is K, S or T.

62. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the amino acid motif comprises the sequence ATK (SEQ ID NO: 1) or ASK (SEQ ID NO: 2).

63. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the amino acid motif comprises the sequence ATKASK (SEQ ID NO: 6).

64. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the amino acid motif comprises the sequence ATKGATK (SEQ ID NO: 7).

65. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the ligand molecule includes an immunoglobulin-binding protein, an enzyme, a target protein binding partner or a nucleic acid molecule.

66. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the ligand molecule is protein A or a domain of protein A.

67. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the fused product is labelled with a detectable label.

68. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that fused product is labelled with a cytotoxic molecule, a fluorescent molecule or radioactive molecule.

69. The method of immobilizing a ligand molecule to a solid support of claim 54, characterized in that the ligand molecule is a protein and said anchor peptide is bound to the N-terminus or the C-terminus of the ligand protein.

70. A method for purifying a molecule of interest in a sample suspected to include the molecule of interest, characterized in that the method comprises:

(a) contacting the sample with a fused product immobilized to a solid support, the fused product comprising a ligand molecule that binds to the molecule of interest and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue, said contacting being under

conditions wherein a complex is formed between the molecule of interest and the ligand molecule if said molecule of interest is present in the sample; and

(b) recovering the molecule of interest from the complex thereby purifying the molecule of interest from the sample.

71. A method for detecting a molecule of interest in a sample suspected to include the molecule of interest, characterized in that the method comprises:

(a) contacting the sample to a fused product immobilized to a solid support, the fused product comprising a ligand molecule that binds to the molecule of interest and an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue, said contacting being under conditions wherein a complex is formed between the molecule of interest and the ligand molecule if said molecule of interest is present in the sample;

(b) recovering the molecule of interest from the complex; and

(c) detecting the recovered molecule of interest.

72. The method for detecting a molecule of interest of claim 71, characterized in that said method further comprises (d) measuring the quantity of recovered molecule of interest.

73. The method for detecting a molecule of interest of claim 71, characterized in that prior to the recovering step, the method further comprises detecting the complex.

74. The methods according to any one of claims 69 and 70, characterized in that said at least one neutral polar residue is serine (S) or threonine (T), and wherein said at least one basic polar residue is lysine (K).

75. The methods according to any one of claims 70 and 71, characterized in that said molecule of interest is IgG and said ligand molecule is protein A or a domain of protein A.

76. A vector comprising at least one copy of a gene of a ligand molecule and at least one gene capable of coding for an anchor peptide, said anchor peptide comprising an amino acid motif having at least one neutral polar residue and at least one basic polar residue.

I.D.	Structure	Sequence
1	X-A-X-G-A-X-G-X	biotin-G-G-Y-G-K-A-K-G-A-K-G-K-A-G
2		biotin-G-G-Y-G-K-A-S-G-A-K-G-S-A-G
3		biotin-G-G-Y-G-S-A-S-G-A-K-G-K-A-G
4		biotin-G-G-Y-G-K-A-S-G-A-K-G-S-A-G
5		biotin-G-G-Y-G-S-A-S-G-A-S-G-S-A-G
6		biotin-G-G-Y-G-T-A-T-G-A-T-G-T-A-G
7		biotin-G-G-Y-G-K-A-T-G-A-K-G-T-A-G
8		biotin-G-G-Y-G-H-A-K-G-A-H-G-K-A-G
9	X-X-G-A-X-X	biotin-G-G-Y-G-A-K-K-G-A-K-K-G-A-G
10		biotin-G-G-Y-G-A-K-S-G-A-K-S-G-A-G
11		biotin-G-G-Y-G-A-S-K-G-A-S-K-G-A-G
12		biotin-G-G-Y-G-A-K-T-G-A-K-T-G-A-G
13		biotin-G-G-Y-G-A-T-K-G-A-T-K-G-A-G
14		biotin-G-G-Y-G-A-S-T-G-A-S-T-G-A-G
15		biotin-G-G-Y-G-A-H-K-G-A-H-K-G-A-G
16		biotin-G-G-Y-G-A-H-H-G-A-K-K-G-A-G
17	X-X-X-G-X-X-X	biotin-G-G-Y-G-A-K-K-K-G-K-K-K-G-A
18		biotin-G-G-Y-G-A-K-S-K-G-K-S-K-G-A
19		biotin-G-G-Y-G-A-K-T-K-G-K-T-K-G-A
20		biotin-G-G-Y-G-A-K-H-K-G-K-H-K-G-A
21		biotin-G-G-Y-G-A-K-K-S-G-K-K-S-G-A
22		biotin-G-G-Y-G-A-K-K-T-G-K-K-T-G-A
23		biotin-G-G-Y-G-A-S-K-K-G-S-K-K-G-A
24		biotin-G-G-Y-G-A-T-K-K-G-T-K-K-G-A
25	Random	biotin-G-K-D-Y-D-I-P-K-K-E-N-L-Y-F-Q
26		biotin-G-H-H-Y-D-I-P-K-K-E-N-L-Y-F-Q
27		biotin-G-S-K-Y-K-I-P-T-T-K-K-L-Y-F-Q
28		biotin-G-S-S-Y-S-A-P-T-T-S-N-L-Y-F-Q
29		biotin-G-S-D-Y-D-A-K-T-T-E-N-L-Y-F-Q
30		biotin-G-S-D-Y-D-A-P-T-T-E-K-K-Y-F-Q
31		biotin-G-S-D-Y-D-A-P-T-T-E-K-K-Y-K-K
32		biotin-G-S-D-Y-D-A-P-T-T-E-K-K-T-K-K
33		biotin-G-S-D-Y-D-A-P-T-T-E-K-K-H-K-K

FIGURE 1

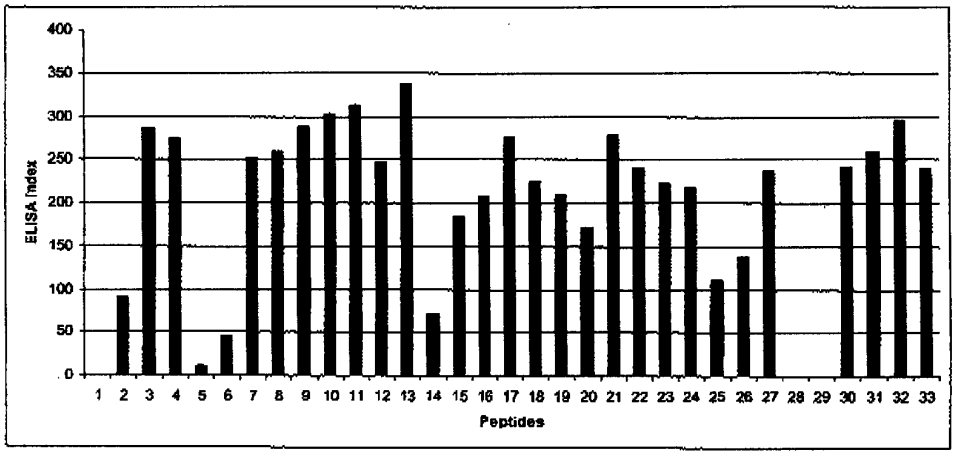

400
350
300
250
200
150
100
50
0
ELISA Index
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
Peptides

FIGURE 2

A


ATKASK-
Z domain

B

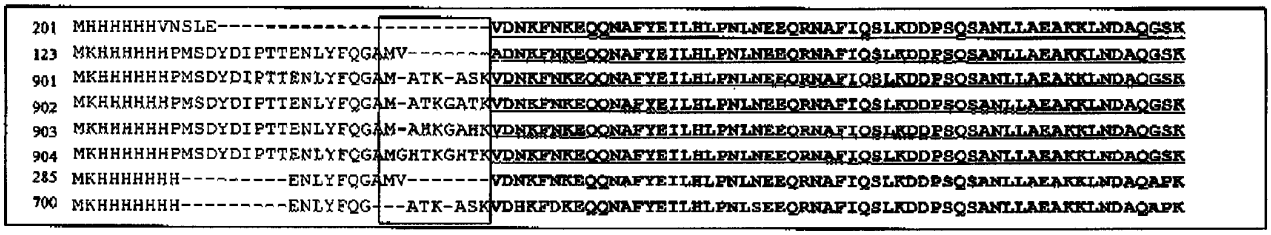

201 MHHHHHHVNSLE----------VDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQGSK
123 MKHHHHHHPMSDYDIPTTENLYFQGAMV-------ADNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQGSK
901 MKHHHHHHPMSDYDIPTTENLYFQGAM-ATK-ASKVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQGSK
902 MKHHHHHHPMSDYDIPTTENLYFQGAM-ATKGATKVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQGSK
903 MKHHHHHHPMSDYDIPTTENLYFQGAM-AHKGAHKVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQGSK
904 MKHHHHHHPMSDYDIPTTENLYFQGAMGHTKGHTKVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQGSK
285 MKHHHHHHH----------ENLYFQGAMV-------VDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQAPK
700 MKHHHHHHH----------ENLYFQG--ATK-ASKVDHKFDKEQQNAFYETLHLPNLSEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQAPK

FIGURE 3

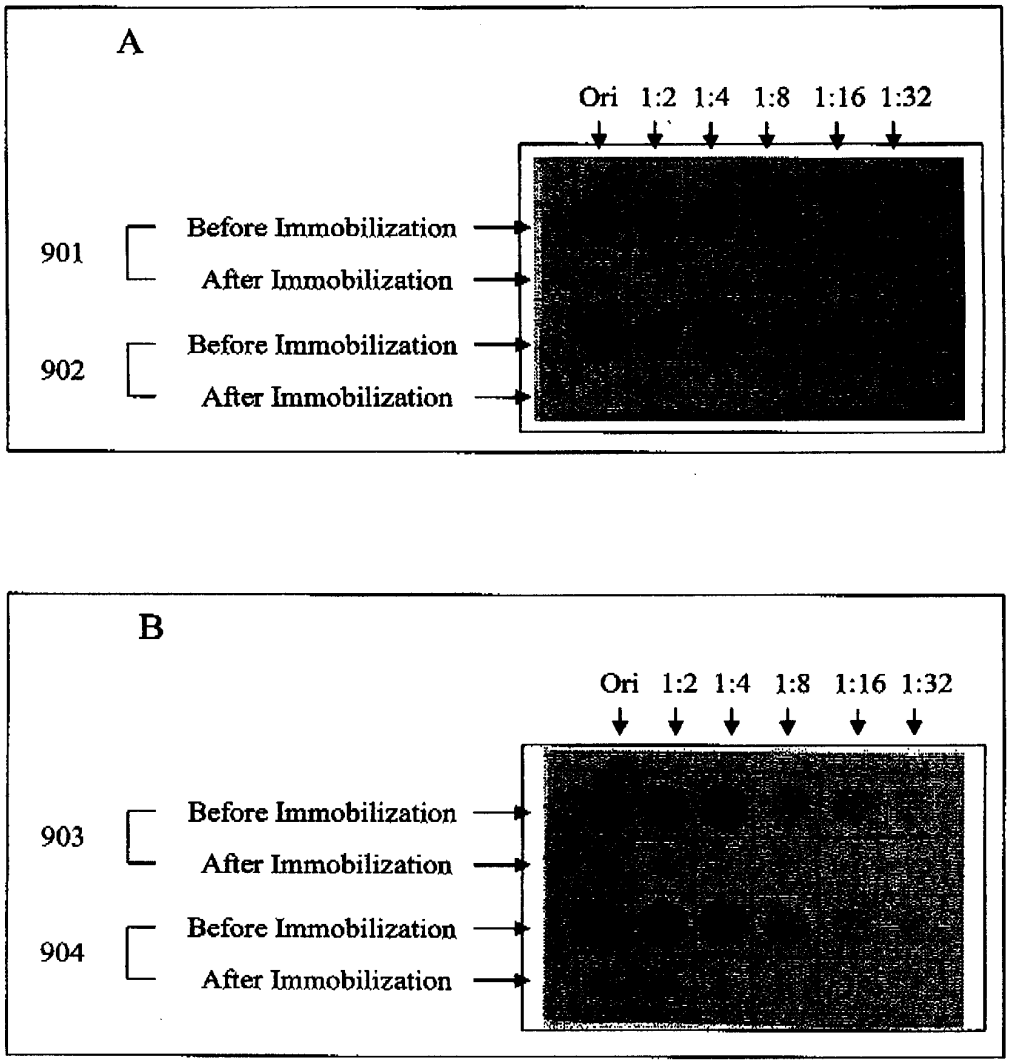

A
Ori 1:2 1:4 1:8 1:16 1:32
901
Before Immobilization
After Immobilization
902
Before Immobilization
After Immobilization
B
Ori 1:2 1:4 1:8 1:16 1:32
903
Before Immobilization
After Immobilization
904
Before Immobilization
After Immobilization

FIGURE 4

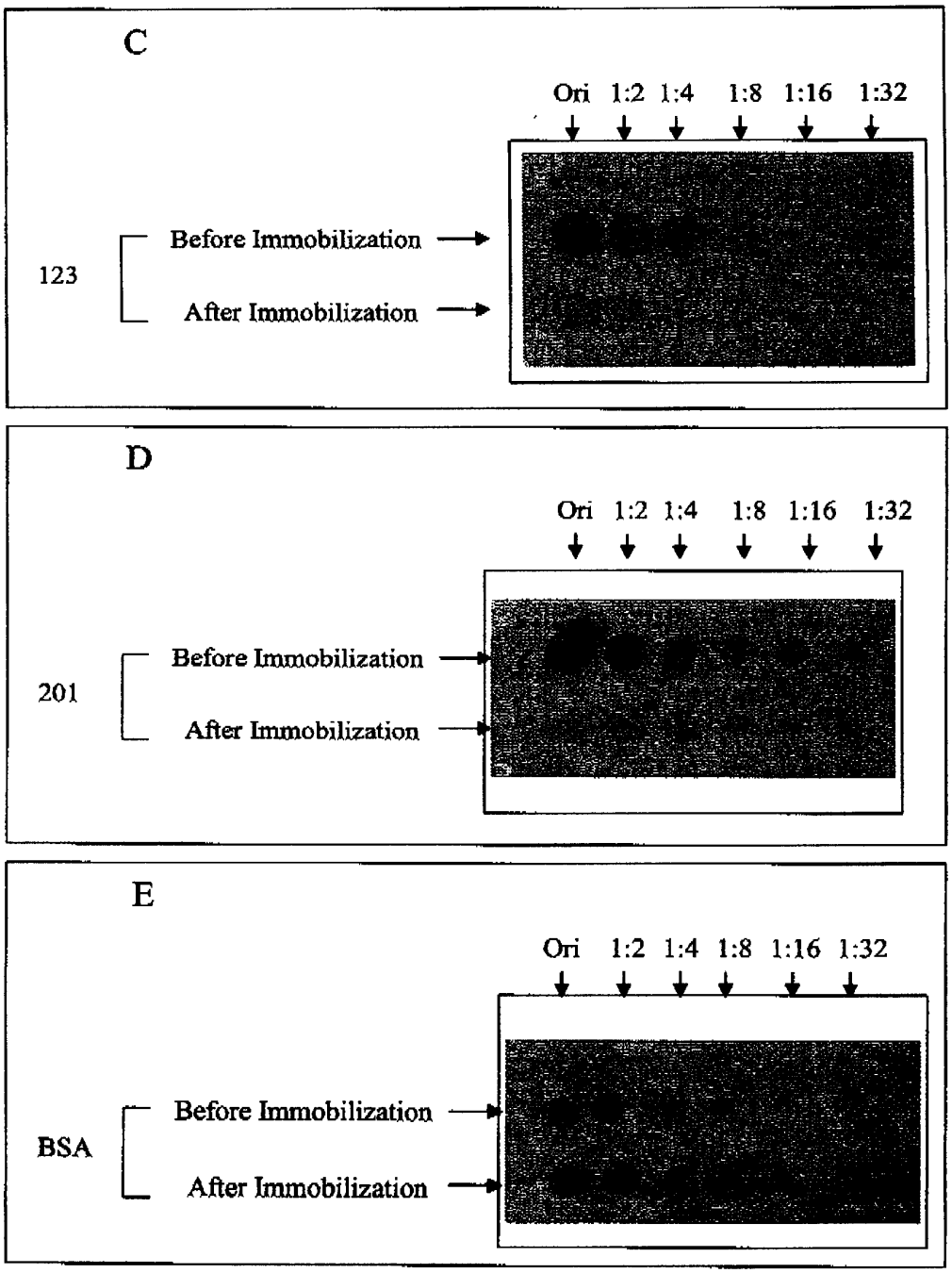


FIGURE 4 (Continued)

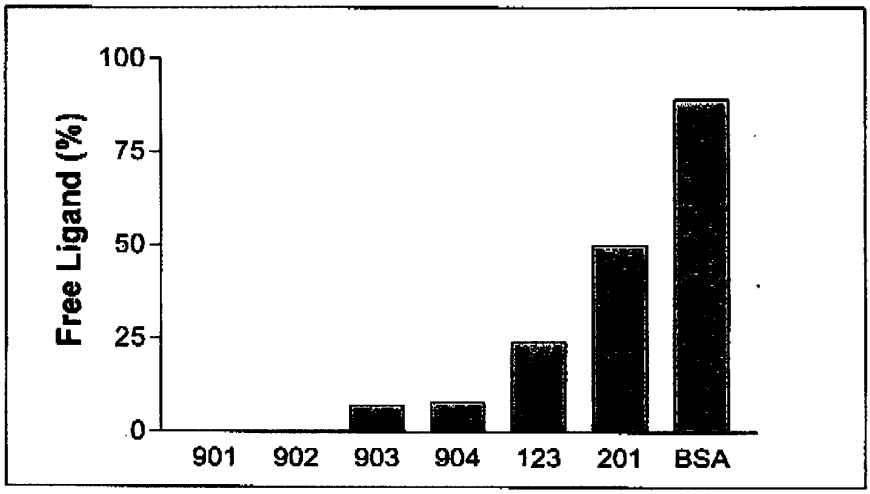
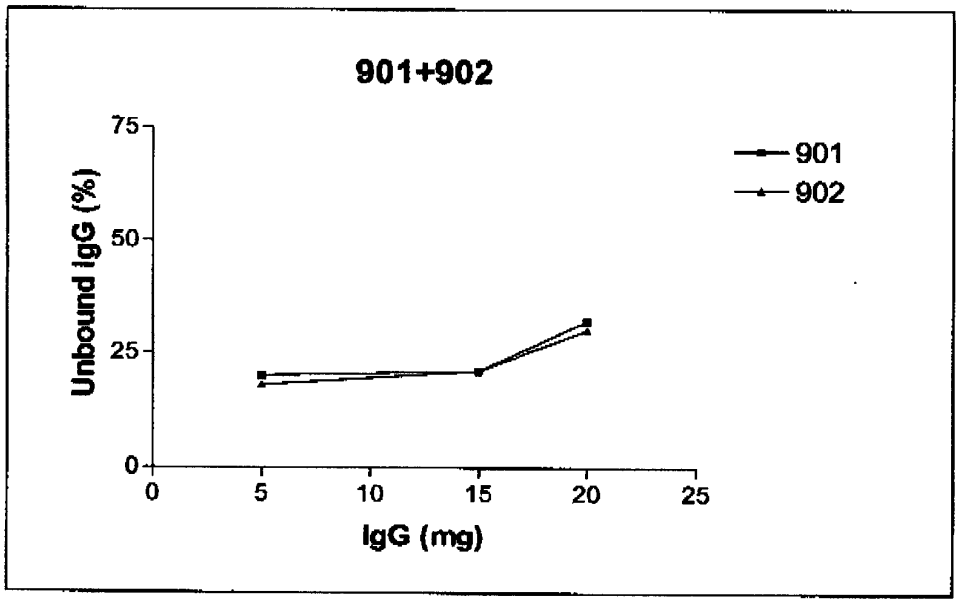


FIGURE 5

A.


901+902
Unbound IgG (%)
75
50
25
0
0
5
10
15
20
25
IgG (mg)
901
902

B.

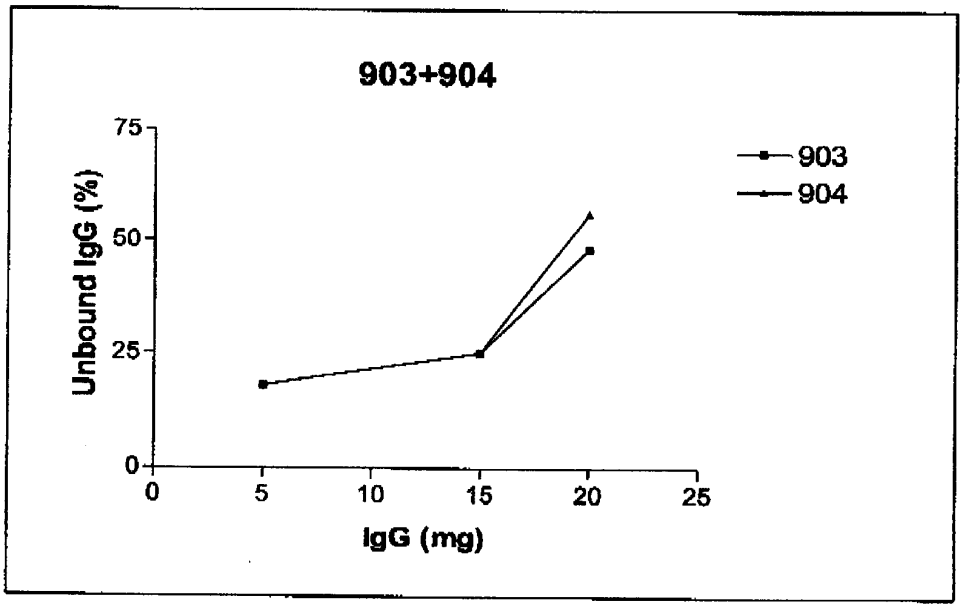
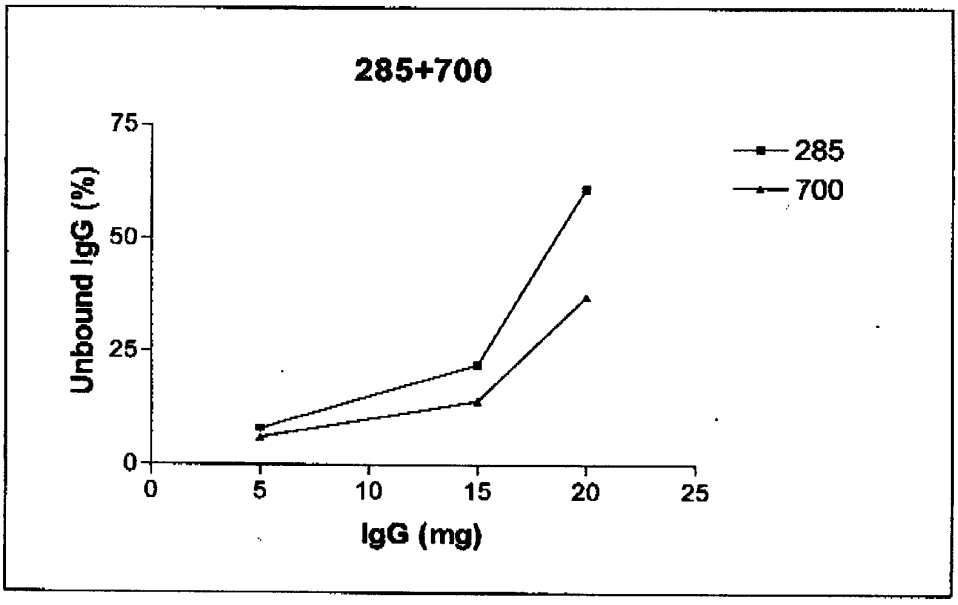

903+904
Unbound IgG (%)
75
50
25
0
0
5
10
15
20
25
IgG (mg)
903
904

FIGURE 6

C.



D.

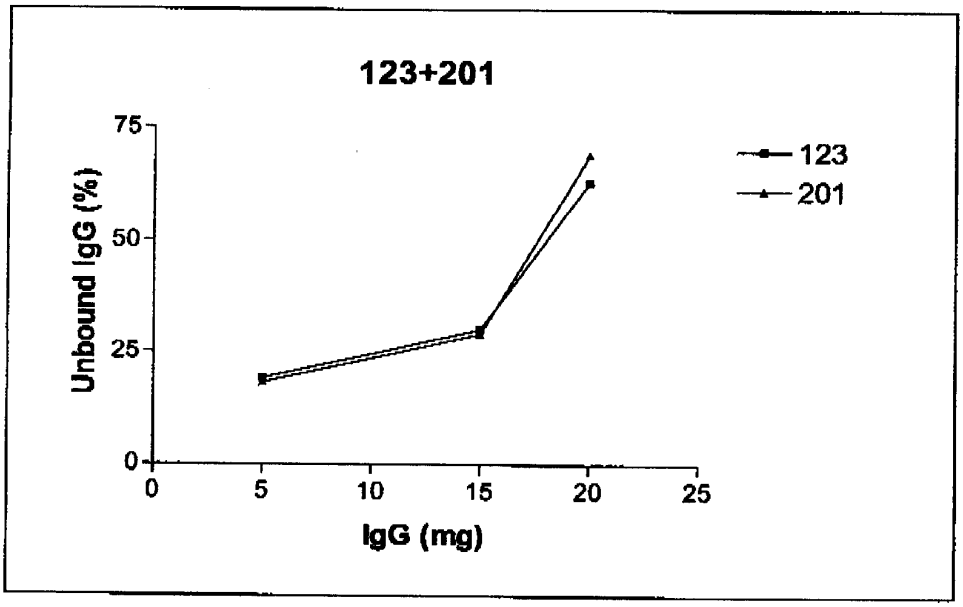


FIGURE 6 (Continued)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2011/001366

A. CLASSIFICATION OF SUBJECT MATTER
IPC: ***C07K 17/00*** (2006.01), ***B01D 15/08*** (2006.01), ***C07K 17/06*** (2006.01), ***C07K 5/083*** (2006.01), ***C07K 7/06*** (2006.01), ***C12N 15/63*** (2006.01) (more IPCs on the last page)
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC: C07K, C12N, B01D, C40B, G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)
Databases: PubMed, Genome Quest, Canadian Patent Database, Scopus, TotalPatent. Keywords: dilysine, serine, threonine, glycine, alanine, fusion, anchor, linker, spacer, epoxy, immobil*, serine, coupl*, conjugate, solid support, substrate, ATK, GASK, ATKGATK,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KUMADA, Y. et al, Development of a one-step ELISA method using an affinity peptide tag specific to a hydrophilic polystyrene surface. J. Biotechnol. 2007, Vol.127, No.2, pages 288-299, ISSN:0168-1656, the whole document.	1, 2, 4-6, 13-18, 20-22, 29-35, 37-39, 46-50, 54-56, 58, 65-76
X --- Y	ALLARD, L. et al, Versatile method for production and controlled polymer-immobilization of biologically active recombinant proteins. Biotechnol. Bioeng. 2002, Vol.80, No.3, pages 341-348, ISSN: 0006-3592, the whole document.	1-4, 6-8, 13-20, 22-24, 29-37, 39-41, 46-56, 58-60, 65-76 --- 9-12, 25-28, 42-45, 61-64
X	ZHU, J. et al, Design and synthesis of biomimetic hydrogel scaffolds with controlled organization of cyclic RGD peptides. Bioconjug. Chem. 2009, Vol.20, No.2, pages 333-339, ISSN:1043-1802, the whole document.	1-3, 6, 13-19, 22, 29-36, 39, 46-55, 58, 65-76

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents :
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 April 2012 (16-04-2012)	10 May 2012 (10-05-2012)
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476	Authorized officer Riad Qanbar (819) 934-7937

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2011/001366

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:

   a. (means)

      [X] on paper

      [X] in electronic form

   b. (time)

      [X] in the international application as filed

      [ ] together with the international application in electronic form

      [ ] subsequently to this Authority for the purposes of search

2. [X] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments :

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2011/001366

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of the first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons :

1. [ ] Claim Nos. :
because they relate to subject matter not required to be searched by this Authority, namely :

2. [ ] Claim Nos. :
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically :

3. [ ] Claim Nos. :
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows :

The claims are directed to a plurality of inventive concepts as follows:

Group 1 - Claims 1, 7-12, 17, 19, 23-28, 34, 36, 40-45, 54, 59-64 (partially) and claims 2, 4-6, 13-16, 18, 20-22, 29-33, 35, 37-39, 46-53, 55-58, 65-76 (completely) featuring an anchor peptide comprising one "neutral polar" residue and one basic polar residue.

Continued in supplementary box

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [X] As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claim Nos. :

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim Nos. :

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

[ ] The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2011/001366

Continuation of Box No. III (Unity of invention is lacking)

Group 2 - Claims 1, 3, 19, 36 (all partially) featuring an anchor peptide of the formula XX, wherein one X is K and the other is S or T.
Group 3 - Claims 1, 7, 23, 40, 59 (all partially) featuring an anchor peptide of the formula XX-UU-XX wherein one X is K, S, or T and U is any amino acid.
Group 4 - Claims 1, 8, 24, 41, 60 (all partially) featuring an anchor peptide of the formula XX-UU-XX wherein one X is A, K, S, or T and U is any amino acid.
Group 5 - Claims 1, 9, 25, 42, 61 (partially) featuring an anchor peptide of the formula XX-GA-XX wherein one X is K, S, or T .
Group 6 - Claims 1, 10, 26, 43, 62 (all partially) featuring an anchor peptide comprising the sequence ATK.
Group 7 - Claims 1, 10, 26, 43, 62 (all partially) featuring an anchor peptide comprising the sequence ASK.
Group 8 - Claims 1, 11, 27, 44, 63 (all partially) featuring an anchor peptide comprising the sequence ATKASK.
Group 9 - Claims 1, 12, 28, 45, 64 (all partially) featuring an anchor peptide comprising the sequence ATKGATK.

The subject matter of groups 1-9 relates to peptides that are useful for linking a protein to a solid substrate. The structural feature that links these groups is the presence of the amino acid lysine in combination with either serine or threonine. This feature, however, was known in the art on the relevant date (see references cited on pages 4 and 5 above).

Additionally, the subject matter of each of groups 1 to 5 comprises multiple sequences that do not have a novel and inventive feature in common and therefore could be further divided into subgroups, each corresponding to a specific, fully defined sequence.

The claims must be limited to one inventive concept as set out in Rule 13 of the PCT.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2011/001366

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 2008/150101 A2 (GIM, Y.S. et al) 11 December 2008 (11-12-2008), the whole document, particularly SEQ ID NOS: 27-30 and 32.	1, 2, 4-6, 13-19, 22, 29-36, 39, 46-55, 58, 65-76 --- 7-12, 25-28, 42-45, 61-64
X	BURZIO, V.A. et al, Mussel adhesive enhances the immobilization of human chorionic gonadotrophin to a solid support. Anal. Biochem. 1996, Vol.241, No.2, pages 190-194, ISSN: 0003-2697, the whole document	1-5, 13
X --- Y	BURZIO, L.A. et al, The adhesive protein of *Choromytilus chorus* (Molina, 1782) and *Aulacomya ater* (Molina, 1782): a proline-rich and a glycine-rich polyphenolic protein. Biochim. Biophys. Acta. 2000, Vol.1479, No.1-2, pages 315-320, ISSN: 0006-3002, the whole document, particularly the third full paragraph of the Discussion.	1-6 --- 7-12, 25-28, 42-45, 61-64
X --- Y	RZEPECKI, L.M. et al, Molecular diversity of marine glues: polyphenolic proteins from five mussel species. Mol. Mar. Biol. Biotechnol. 1991, Vol.1, No.1, pages 78-88, ISSN: 1053-6426, the whole document.	1-8 --- 7-12, 25-28, 42-45, 61-64
X	MATSUNAGA, T. et al, Hypothetical protein [*Magnetospirillum magneticum AMB-1*]. GenBank, Accession Number: BAE51773, first seen at NCBI on 1 December 2005 (01-12-2010).	1-11
X	MARDANI, A. et al, Knob-associated histidine-rich protein, partial [*Plasmodium falciparum*]. GenBank, Accession Number: ADM31373, first seen at NCBI on 29 August 2010 (29-08-2010).	1-10, 12
A	MATEO, C. et al, Advances in the design of new epoxy supports for enzyme immobilization-stabilization. Biochem. Soc. Trans. 2007, Vol.35, No.6, pages1593-1601, ISSN: 0300-5127.	1-76

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CA2011/001366

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
WO2008150101A2	11 December 2008 (11-12-2008)	WO2008150101A2	11 December 2008 (11-12-2008)
		WO2008150101A3	29 January 2009 (29-01-2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2011/001366

C40B 30/04 (2006.01), ***C40B 50/18*** (2006.01), ***C40B 80/00*** (2006.01), ***G01N 33/53*** (2006.01), ***G01N 33/564*** (2006.01), *C07K 1/22* (2006.01)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2013 (17.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/010042 A1

(51) International Patent Classification:
A23J 1/00 (2006.01)

(21) International Application Number:
PCT/US2012/046560

(22) International Filing Date:
12 July 2012 (12.07.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/572,205 12 July 2011 (12.07.2011) US

(71) Applicant *(for all designated States except US)*: **SAND HILL FOODS. INC.** [US/US]; 1140 O'brien Drive, Suite B, Menlo Park, CA 94025 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BROWN, Patrick** [US/US]; 76 Peter Coutts Circle, Stanford, CA 94305 (US). **VRLJIC, Marija** [US/US]; 130 Barneson Avenue, Apt. 4, San Mateo CA 94402 (US). **VARADAN, Ranjani** [IN/US]; 20 Descanso Drive, Unit 1311, San Jose, CA 95134 (US). **EISEN, Michael** [US/US]; 2408 McGee Avenue, Berkeley, CA 94703 (US). **SOLOMATIN, Sergey** [RU/US]; 422 Hillwood Ct, Mountain View, CA 94040 (US).

(74) Agents: **WILLIS, Michael, A.** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2013/010042 A1

(54) **Title**: METHODS AND COMPOSITIONS FOR CONSUMABLES

(57) **Abstract**: Provided herein are methods and compositions related to plant based meat substitutes which have properties similar to meat.

METHODS AND COMPOSITIONS FOR CONSUMABLES

This application claims priority to U.S. application Ser. No. 61/572,205 filed July 12, 2011, which is incorporated herein in its entirety by reference.

BACKGROUND OF THE INVENTION

[0001] Animal farming has a profound negative environmental impact. Currently it is estimated that 30% of Earth's land surface is dedicated to animal farming and that livestock account for 20% of total terrestrial animal biomass. Due to this massive scale animal farming accounts for more than 18% of net greenhouse gas emissions. Animal farming may be the largest human source of water pollution, and animal farming is by far the world's largest threat to biodiversity. It has been estimated that if the worlds human population could shift from a meat containing diet to a diet free of animal products, 26% of Earth's land surface would be freed for other uses. Furthermore the shift to a vegetarian diet would massively reduce water and energy consumption.

[0002] The consumption of meat has a profound negative impact on human health. The heath benefits of a vegetarian diet are well established. If the human population would shift to a vegetarian diet the cost savings in health care would be significant.

[0003] Hunger is a worldwide problem, yet the world's 4 major commodity crops (soybeans, maize, wheat, and rice) already supply more than 100% of the human population's requirements for calories and protein, including every essential amino acid.

[0004] Plant based meat substitutes have largely failed to cause a shift to a vegetarian diet. The current state of the art for meat substitute compositions involves the extrusion of soy/grain mixture, resulting in products which largely fail to replicate the experience of cooking and eating meat. Common limitations of these products are a texture and mouthfeel that are more homogenous than that of equivalent meat products. Furthermore, as the products must largely be sold pre-cooked, with artificial flavors and aromas built in, they fail to replicate aromas, flavors, and other key features associated with cooking meat. As a result, these products appeal mainly to a limited consumer base that is already committed to vegetarianism/veganism, but have failed to appeal to the larger consumer segment accustomed to eating meat.

[0005] Disclosed herein are improved methods and compositions which more accurately replicate the characteristics that consumers value in the preparation and consumption of meat and which overcome the shortcomings and drawbacks of current meat substitutes.

SUMMARY OF THE INVENTION

[0006] In some aspects, the invention provides a meat substitute composition comprising a protein content, wherein one or more isolated and purified proteins accounts for 10% or more of said protein content by weight, wherein said meat substitute composition accurately mimics the taste, texture, or color of a meat product derived from animal sources.

[0007] In one embodiment, the meat substitute composition accurately mimics the color of said meat product in its raw state and in a cooked state after cooking.

[0008] In another embodiment, the one or more isolated and purified proteins accounts for 25% or more of said protein content by weight.

[0009] In another embodiment, the one or more isolated and purified proteins accounts for 50% or more of said protein content by weight.

[0010] In another embodiment, the one or more isolated and purified proteins accounts for 75% or more of said protein content by weight.

[0011] In another embodiment, the one or more isolated and purified proteins accounts for 90% or more of said protein content by weight.

[0012] In another embodiment, gluten does not account for 10% or more of said protein content by weight.

[0013] In another embodiment, each of said isolated, purified proteins is separately isolated and purified.

[0014] In another embodiment, the meat substitute composition comprises 1-7 isolated and purified proteins.

[0015] In another embodiment, said 1-7 isolated and purified proteins are each isolated from different plant species.

[0016] In some embodiments the meat substitute comprises less than 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 individual proteins isolated from one or more plant species.

[0017] In another embodiment, said protein content comprises no more than trace amounts of any other proteins derived from the one or more plant species.

[0018] In another embodiment, said one or more isolated and purified proteins are selected from the group consisting of leghemoglobin, non-symbiotic hemoglobin, hemoglobin, myoglobin, chlorocruorin, erythrocruorin, neuroglobin, cytoglobin, protoglobin, truncated 2/2 globin, HbN, cyanoglobin, HbO, Glb3, and cytochromes, Hell's gate globin I, bacterial hemoglobins, ciliate myoglobins, flavohemoglobins, ribosomal proteins, actin, hexokinase, lactate dehydrogenase, fructose bisphosphate aldolase, phosphofructokinases, triose phosphate isomerases, phosphoglycerate kinases, phosphoglycerate mutases, enolases, pyruvate kinases, glyceraldehyde-3-phosphate dehydrogenases, pyruvate decarboxylases, actins, translation elongation factors, ribulose-1,5-bisphosphate carboxylase oxygenase (rubisco), ribulose-1,5-bisphosphate carboxylase oxygenase activase (rubisco activase), albumins, glycinins, conglycinins, globulins, vicilins, conalbumin, gliadin, glutelin, gluten, glutenin, hordein, prolamin, phaseolin (protein), proteinoplast, secalin, extensins, triticeae gluten, zein, any seed storage protein, oleosins, caloleosins, steroleosins or other oil body proteins, vegetative storage protein A, vegetative storage protein B, moong seed storage 8S globulin.

[0019] In another embodiment, said one or more isolated and purified proteins are not isolated from an animal.

[0020] In another embodiment, said one or more isolated and purified proteins are isolated from a single plant source.