

(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
20. Februar 2003 (20.02.2003)

PCT

(10) Internationale Veröffentlichungsnummer
WO 03/014353 A2

(51) Internationale Patentklassifikation⁷: **C12N 15/31**,
C07K 14/37, C12N 15/80, I/15, C12Q 1/02, C12P 19/42

(21) Internationales Aktenzeichen: PCT/EP02/08937

(22) Internationales Anmeldedatum:
9. August 2002 (09.08.2002)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:

101 39 454.3	10. August 2001 (10.08.2001)	DE
101 39 455.1	10. August 2001 (10.08.2001)	DE
101 39 457.8	10. August 2001 (10.08.2001)	DE
101 39 458.6	10. August 2001 (10.08.2001)	DE
101 39 459.4	10. August 2001 (10.08.2001)	DE
101 39 460.8	10. August 2001 (10.08.2001)	DE
101 39 461.6	10. August 2001 (10.08.2001)	DE
101 39 462.4	10. August 2001 (10.08.2001)	DE
101 39 463.2	10. August 2001 (10.08.2001)	DE
101 39 464.0	10. August 2001 (10.08.2001)	DE
102 09 819.0	6. März 2002 (06.03.2002)	DE
102 09 816.6	6. März 2002 (06.03.2002)	DE
102 16 033.3	11. April 2002 (11.04.2002)	DE
102 21 911.7	16. Mai 2002 (16.05.2002)	DE
102 21 928.1	16. Mai 2002 (16.05.2002)	DE
102 21 909.5	16. Mai 2002 (16.05.2002)	DE
102 25 390.0	7. Juni 2002 (07.06.2002)	DE
102 25 392.7	7. Juni 2002 (07.06.2002)	DE
102 27 796.6	21. Juni 2002 (21.06.2002)	DE
102 34 455.8	29. Juli 2002 (29.07.2002)	DE

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme
von US): **BASF AKTIENGESellschaft** [DE/DE];
67056 Ludwigshafen (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): **KAROS, Marvin**
[DE/DE]; Fröbelstrasse 2, 67433 Neustadt (DE). **AL-**
THÖFER, Henning [DE/DE]; Im Höhnhausen 26, 67157
Wachenheim (DE). **KRÖGER, Burkhard** [DE/DE];
Im Waldhof 1, 67117 Limburgerhof (DE). **REVUELTA**
DOVAL, Jose, L. [ES/ES]; Grillo 11, 4E, E-37001 Sala-
manca (ES).

(74) Anwalt: **SCHWEIGER, Georg**; Reitstötter, Kinzebach &
Partner, Sternwartstrasse 4, 81679 München (DE).

(81) Bestimmungsstaaten (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, NO, NZ, OM, PH, PL, PT, RO, RU,
SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VN, YU, ZA, ZM, ZW.

(84) Bestimmungsstaaten (regional): ARIPO-Patent (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, IE, IT, LU, MC, NL, PT,
SE, 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.

(54) Title: NOVEL GENETIC PRODUCTS OF *ASHBYA GOSSYPPII*, ASSOCIATED WITH TRANSMEMBRANE TRANSPORT

WO 03/014353 A2

(54) Bezeichnung: NEUE GENPRODUKTE AUS *ASHBYA GOSSYPPII*, DIE MIT DEM TRANSMEMBRAN-TRANSPORT ASSOZIIERT SIND

(57) Abstract: The invention relates to novel polynucleotides of *Ashbya gossypii*, oligonucleotides which hybridise with the same, expression cassettes and vectors containing said polynucleotides, micro-organisms transformed by the same, polypeptides coded by said polynucleotides, and the use of the novel polypeptides and polynucleotides as targets for improving transmembrane transport, and especially for improving vitamin B2 production in micro-organisms of the *Ashbya* genus.

(57) Zusammenfassung: Die Erfindung betrifft neuartige Polynukleotide aus *Ashbya gossypii*, damit hybridisierende Oligonukleotide; Expressionskassetten und Vektoren, welche diese Polynukleotide enthalten; damit transformierte Mikroorganismen; von diesen Polynukleotiden kodierte Polypeptide; und die Anwendung der neuen Polypeptide und Polynukleotide als Targets zur Verbesserung des Transmembran-Transports und insbesondere der Verbesserung der Vitamin B2-Produktion in Mikroorganismen der Gattung *Ashbya*.

Beschreibung

Neue Genprodukte aus *Ashbya gossypii*, die mit dem Transmembran-Transport assoziiert sind.

- 5 Die vorliegende Erfindung betrifft neuartige Polynukleotide aus *Ashbya gossypii*; damit hybridisierende Oligonukleotide; Expressionskassetten und Vektoren, welche diese Polynukleotide enthalten; damit transformierte Mikroorganismen; von diesen Polynukleotiden kodierte Polypeptide; und die Anwendung der neuen Polypeptide und Polynukleotide als Targets zur Modulation des Transmembran-Transports und insbesondere der Verbesserung der Vitamin B2-Produktion
10 in Mikroorganismen der Gattung *Ashbya*.

Vitamin B2 (Riboflavin, Lactoflavin) ist ein alkali- und lichtempfindliches, in Lösung gelbgrün fluoreszierendes Vitamin. Vitamin B2-Mangel kann zu Ektodermeschäden, insbesondere Linsentrübung, Keratitis, Korneavaskularisation, zu neurovegetativen und urogenitalen Störungen führen.

- 15 Vitamin B2 ist Vorläufer für die neben NAD^+ und NADP^+ wichtigen biologischen Wasserstoffüberträger-Moleküle FAD und FMN. Diese werden aus Vitamin B2 durch Phosphorylierung (FMN) und anschließende Adenylierung (FAD) gebildet.

- Vitamin B2 wird in Pflanzen, Hefen und vielen Mikroorganismen aus GTP und Ribulose-5-phosphat synthetisiert. Der Reaktionsweg beginnt mit dem Öffnen des Imidazolrings von GTP und Abspaltung eines Phosphatrestes. Durch Desaminierung, Reduktion und Abspaltung des verbleibenden Phosphats entsteht 5-Amino-6-ribitylamino-2,4-pyrimidinon. Die Reaktion dieser Verbindung mit 3,4-Dihydroxy-2-butanon-4-phosphat führt zum bicyclischen Molekül 6,7-Dimethyl-8-ribityllumazin. Diese Verbindung wird durch Dismutation, bei der eine 4-Kohlenstoff-
20 Einheit übertragen wird, in die tricyclische Verbindung Riboflavin umgesetzt.

- Vitamin B2 kommt in vielen Gemüsen und in Fleisch vor, weniger in Getreideprodukten. Der tägliche Vitamin B2-Bedarf eines Erwachsenen liegt bei etwa 1,4 bis 2 mg. Hauptabbauprodukt der Coenzyme FMN und FAD beim Menschen ist wiederum Riboflavin, welches als solches ausge-
30 schieden wird.

- Vitamin B2 stellt damit ein wichtiges Nahrungsergänzungsmittel für Mensch und Tier dar. Es besteht daher das Bestreben, Vitamin B2 in technischem Maßstab zugänglich zu machen. Es wurde daher vorgeschlagen, Vitamin B2 auf mikrobiologischem Weg zu synthetisieren. Brauchbare Mikroorganismen hierfür sind beispielsweise *Bacillus subtilis*, die Ascomyceten *Eremothecium ashbyii*, *Ashbya gossypii* sowie die Hefen *Candida flareri* und *Saccharomyces cerevisiae*. Die hierzu verwendeten Nährmedien umfassen Melasse oder Pflanzenöle als Kohlenstoffquelle, anorganische Salze, Aminosäuren, tierische oder pflanzliche Peptone und Proteine sowie Vita-

minzusätze. In sterilen aeroben submersen Verfahren erhält man pro Liter Kulturbrühe Ausbeuten von mehr als 10 g Vitamin B2 innerhalb weniger Tage. Voraussetzung sind gute Belüftung der Kultur, sorgfältiges Rühren und Einstellung von Temperaturen unter etwa 30°C. Nach Abtrennen der Biomasse, Eindampfen und Trocknen des Konzentrates erhält man ein mit Vitamin B2 angereichertes Produkt.

Die mikrobiologische Produktion von Vitamin B2 ist beispielsweise beschrieben in der WO-A-92/01060, der EP-A-0 405 370 und EP-A-0 531 708.

10 Eine Übersicht über Bedeutung, Vorkommen, Herstellung, Biosynthese und Verwendung von Vitamin B2 ist beispielsweise in Ullmann's Encyclopaedia of Industrial Chemistry, Band A27, Seiten 521 ff. zu finden.

15 Die Zellmembranen dienen in einer Zelle einer Reihe von Funktionen. Zuerst differenziert eine Membran den Zellinhalt von der Umgebung, so dass die Zelle Integrität erhält. Die Membranen dienen auch als Schranken, damit gefährliche oder ungewünschte Verbindungen nicht einströmen können und gewünschte Verbindungen nicht ausströmen können. Zellmembranen sind aufgrund ihrer Struktur von Natur aus gegenüber der nicht erleichterten Diffusion hydrophiler Verbindungen, wie Proteine, Wassermolekülen und Ionen undurchlässig: eine Doppelschicht aus Lipidmolekülen, in der die polaren Kopfgruppen nach außen ragen (aus der Zelle heraus bzw. ins Zellinnere hinein) und die unpolaren Schwänze zur Mitte der Doppelschicht ragen und einen hydrophoben Kern bilden (für einen allgemeinen Überblick über die Struktur und Funktion der Membran siehe Gennis, R.B. (1989) Biomembranes, Molecular Structure and Function, Springer: Heidelberg). Diese Schranke ermöglicht, dass die Zellen eine relativ höhere Konzentration an gewünschten Verbindungen und eine relativ kleinere Konzentration an ungewünschten Verbindungen als das umgebende Medium enthalten, da die Diffusion dieser Verbindungen durch die Membran effizient blockiert wird.

30 Die Membran liefert jedoch auch eine wirksame Schranke gegenüber dem Import von gewünschten Molekülen und dem Export von Abfallmolekülen. Zur Bewältigung dieser Schwierigkeit enthalten die Zellmembranen viele Arten von Transporterproteinen, die den Transmembrantransport verschiedenartiger Verbindungen erleichtern können: Poren oder Kanäle und Transporter. Die ersteren sind integrale Membranproteine, gelegentlich Proteinkomplexe, die eine regulierte Öffnung durch die Membran bilden. Diese Regulation oder dieses "Gating" ist gewöhnlich für die durch die Pore oder den Kanal zu transportierenden Substrate spezifisch, so dass diese Transmembrankonstrukte für eine spezifische Klasse von Substraten spezifisch sind; beispielsweise ist ein Kaliumkanal derart konstruiert, dass nur Ionen mit ähnlicher Ladung und

Größe wie Kalium hindurchgelangen können. Kanal- und Porenproteine haben bestimmte hydrophobe und hydrophile Domänen, so dass sich der hydrophobe Anteil des Proteins an das Innere der Membran anlagern kann, wohingegen der hydrophile Anteil das Innere des Kanals ausmacht, wodurch eine geschützte hydrophile Umgebung bereitgestellt wird, durch die das ausgewählte hydrophile Molekül gelangen kann. Im Fachgebiet sind viele solche Poren/Kanäle bekannt, einschließlich solche für Kalium-, Calcium-, Natrium- und Chloridionen.

Dieses durch Poren und Kanäle vermittelte System ist auf sehr kleine Moleküle, wie Ionen, eingeschränkt, da Poren oder Kanäle, die hinreichend groß sind, dass sie den Durchtritt vollständiger Proteine durch erleichterte Diffusion ermöglichen, auch nicht dazu fähig wären, den Durchtritt kleinerer Moleküle zu verhindern. Der Transport von Molekülen durch diesen Prozeß wird gelegentlich als "erleichterte Diffusion" bezeichnet, da die treibende Kraft eines Konzentrationsgradienten erforderlich ist, damit der Transport stattfindet. Permeasen ermöglichen ebenfalls die erleichterte Diffusion größerer Moleküle, wie Glucose oder anderer Zucker, in die Zelle, wenn die Konzentration dieser Moleküle auf einer Seite der Membran größer ist als auf der anderen (ebenfalls als "Uniport" bezeichnet). Im Gegensatz zu Poren oder Kanälen bilden diese integralen Proteine (die oft 6 bis 14 membranüberspannende Helices aufweisen) keine offenen Kanäle durch die Membran, sie binden jedoch an das Zielmolekül an der Membranoberfläche und durchlaufen dann eine Konformationsänderung, so dass das Zielmolekül an der entgegengesetzten Seite der Membran freigesetzt wird.

Zellen benötigen jedoch oft den Import oder Export von Molekülen gegen den bestehenden Konzentrationsgradienten ("aktiver Transport"), eine Situation, in der die erleichterte Diffusion nicht stattfinden kann. Es gibt zwei generelle Mechanismen, die von der Zelle für einen solchen Membrantransport verwendet werden; Symport oder Antiport, und energiegekoppelter Transport, wie derjenige, der durch ABC-Transporter vermittelt wird. Symport- und Antiportsysteme koppeln die Bewegung von zwei unterschiedlichen Molekülen über die Membran (über Permeasen mit zwei gesonderten Bindungsstellen für zwei unterschiedliche Moleküle); beim Symport werden beide Moleküle in die gleiche Richtung transportiert, wohingegen beim Antiport ein Molekül importiert und das andere Molekül exportiert wird. Dies ist energetisch möglich, da sich eines dieser beiden Moleküle entlang eines Konzentrationsgradienten bewegt, und dieses energetisch günstige Ereignis wird nur durch die gleichzeitige Bewegung einer gewünschten Verbindung gegen den herrschenden Konzentrationsgradienten ermöglicht. Einzelne Moleküle können gegen den Konzentrationsgradienten über die Membran in einem energiegetriebenen Prozeß transportiert werden, wie bspw. bei den ABC-Transportern. Bei diesem System hat das in der Membran lokalisierte Transportprotein eine ATP-bindende Cassette, beim Binden des Zielmoleküls wird ATP in ADP + Pi umgewandelt, und die resultierende freigesetzte Energie wird zum

Antreiben der Bewegung des Zielmoleküls zur entgegengesetzten Seite der Membran verwendet, was durch den Transporter erleichtert wird. Für eingehendere Beschreibungen all dieser Transportsysteme, siehe Bamberg, E. et al., (1993) "Charge transport of ion pumps on lipid bilayer membranes", Q. Rev. Biophys. 26: 1-25; Findlay, J.B.C. (1991) "Structure and function in membrane transport systems", Curr. Opin. Struct. Biol. 1: 804-810; Higgins, C.F. (1992) "ABC transporters from microorganisms to man", Ann. Rev. Cell. Biol. 8: 67-113; Gennis, R.B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, S. 270-322; und Nikaido, H. und Saier, H. (1992) "Transport proteins in bacteria: common themes in their design", Science 258: 936-942, und die in jeder dieser Zitate enthaltenen Literaturstellen. Die Nutzung von Genen des Transmembran-Transports zur Generierung von Mikroorganismen, vorzugsweise der Gattung *Ashbya*, insbesondere von *Ashbya gossypii* Stämmen, mit veränderten Transmembran-Transporteigenschaften ist noch nicht beschrieben.

Aufgabe der vorliegenden Erfindung ist daher die Bereitstellung neuer Targets zur Beeinflussung des Transmembran-Transports in Mikroorganismen der Gattung *Ashbya*, insbesondere in *Ashbya gossypii*. Insbesondere besteht die Aufgabe, den Transmembran-Transport in derartigen Mikroorganismen zu verbessern. Eine weitere Aufgabe ist die Verbesserung der Vitamin B2-Produktion durch derartige Mikroorganismen.

Gelöst wird obige Aufgabe insbesondere durch Bereitstellung kodierender Nukleinsäuresequenzen, welche in *Ashbya gossypii* während der Vitamin B2-Produktion hoch- bzw. niederreguliert (basierend auf Ergebnissen, ermittelt mit Hilfe der im experimentellen Teil näher beschriebenen MPSS-Analysenmethode) sind.

Gelöst wird obige Aufgabe insbesondere durch Bereitstellung von Polynukleotiden, welche aus *Ashbya gossypii* isolierbar sind und für ein Protein kodieren, das mit dem Transmembrantransport assoziiert ist und/oder ein Transmembranprotein ist und insbesondere eine in Tabelle 1 angegebene strukturelle (z.B. Sequenzhomologie) und/oder funktionale Eigenschaft (z.B. Enzymaktivität) besitzen; und zwar insbesondere:

a) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion eines mitochondrialen energieübertragenden Proteins kodiert.

Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 19“ trägt.

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 19v“ trägt.

5

Ein Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 1. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 3 oder ein Fragment davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

Die Inserts von „Oligo 19“ und „Oligo 19v“ besitzen mit dem MIPS Tag „Ygr257c“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 1 bzw. SEQ ID NO: 3. Die vom korrespondierenden Gegenstrang von SEQ ID NO: 1 bzw. von kodierenden Strang gemäß SEQ ID NO: 3 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem mitochondrialen energieübertragenden Protein aus *S. cerevisiae*.

20

b) eine, vorzugsweise hochregulierte Nukleinsäuresequenz, welche für ein Protein mit der Funktion entsprechend der eines ABC-Transportproteins aus *S. cerevisiae* kodiert. ABC (ATP binding cassette) Proteine fungieren als Transportsysteme und sind an der Aufnahme oder der Abgabe von Substraten aus der Zelle beteiligt. Der Transportvorgang wird dabei durch ATP-Hydrolyse getrieben.

25

Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 24“ trägt.

30

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 24“ trägt.

Ein Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 5. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 8 oder ein Fragment da-

35

von. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

5

Die Inserts von „Oligo 24“ und „Oligo 24v“ besitzen mit dem MIPS Tag „Mdl2“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 5 bzw. SEQ ID NO: 8. Die vom korrespondierenden Gegenstrang zu SEQ ID NO:5 bzw. von kodierenden Strang gemäß SEQ ID NO:8 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem ABC-Transportprotein aus *S. cerevisiae*.

10

c) eine, vorzugsweise niederregulierte Nukleinsäuresequenz, welche für ein Protein mit der Funktion eines membranintegrierten mitochondrialen Proteins kodiert.

15

Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 109“ trägt.

20

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 109v“ trägt.

25

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 10. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 12 oder ein Fragment davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

30

Die Inserts von „Oligo 109“ und „Oligo 109v“ besitzen mit dem MIPS Tag „Prp12“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 10 bzw. SEQ ID NO: 12. Die vom kodierenden Strang abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem membranintegrierten mitochondrialen Protein aus *S. cerevisiae*.

35

d) eine, vorzugsweise niederregulierte Nukleinsäuresequenz, welche für ein Protein mit der Funktion eines mitochondrialen Innermembran-Transportproteins kodiert.

5 Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein cDNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 163“ trägt.

10 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, welcher für die Vollsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 163v“ trägt.

15 Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 14. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 17 oder Fragmente davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar ist. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

20 Die Inserts von „Oligo 163“ und „Oligo 163v“ besitzen mit dem MIPS Tag „Flx1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 14 bzw. SEQ ID NO: 17. Die vom kodierenden Strang abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem mitochondrialen Innermembran-Transportprotein aus *S. cerevisiae*.

25

e) eine, vorzugsweise niederregulierte Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer nicht-vakuolären 102 kD-Untereinheit der H⁺-ATPase-VO-Domäne kodiert.

30 Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 31“ trägt.

35 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 31v“ trägt.

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 19. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 21 oder ein Fragment davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

Die Inserts von „Oligo 31“ und „Oligo 31v“ besitzen mit dem MIPS Tag „STV1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 19 bzw. SEQ ID NO: 21. Die vom kodierenden Strang abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer nicht-vakuolären 102 kD-Untereinheit der H⁺-ATPase-VO-Domäne aus *S. cerevisiae*.

f) eine, vorzugsweise hochregulierte Nukleinsäuresequenz, welche für ein Protein mit einer Funktion, die eine Ähnlichkeit zu der des isp4-Proteins aus *S. pombe* aufweist, kodiert.

Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde erfindungsgemäß ein cDNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 4“ trägt.

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 4v“ trägt.

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 23. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 26 oder die dazu komplementäre Sequenz gemäß SEQ ID NO: 25. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar ist. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

Die Inserts von „Oligo 4“ und „Oligo 4v“ besitzen mit dem MIPS Tag „OPT2“ aus *S. cerevisiae* signifikante Homologien. Das Insert umfasst eine Nukleinsäuresequenz gemäß SEQ ID NO: 23 bzw. 25. Die vom kodierenden Strang (umfassend SEQ ID NO: 26) abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem Protein

aus *S. cerevisiae* mit einer Ähnlichkeit zum isp4-Protein aus *S. pombe*. Den erfindungsgemäßen Proteinen wird daher die Aktivität eines Oligopeptid-Transporters zugeordnet.

- 5 g) eine, vorzugsweise hochregulierte Nukleinsäuresequenz, welche für ein Protein mit der Funktion eines VAC1-Proteins aus *S. cerevisiae*, einem cytosolischen und peripheren Membranprotein mit drei Zinkfingern, kodiert.

10 Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 6“ trägt.

15 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 6v“ trägt.

20 Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 28. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 31 oder ein Fragment davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

25 Die Inserts von „Oligo 6“ und „Oligo 6v“ besitzen mit dem MIPS Tag „VAC1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 28 bzw. SEQ ID NO: 31. Die vom korrespondierenden Gegenstrang zu SEQ ID NO:28 bzw. vom Strang gemäß SEQ ID NO:31 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem VAC1-Protein, einem cytosolischen und peripheren Membranprotein mit drei Zinkfingern, aus *S. cerevisiae*.

- 30 h) eine, vorzugsweise hochregulierte Nukleinsäuresequenz, welche für ein Protein mit einer ATPase-ähnlichen Funktion kodiert.

35 Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 146“ trägt.

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 146v“ trägt.

- 5 Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 33. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 35 oder ein Fragment davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu kom-
10 plementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

- Die Inserts von „Oligo 146“ und „Oligo 146v“ besitzen mit dem MIPS Tag „Ymr162c“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID
15 NO: 33 bzw. SEQ ID NO: 35. Die vom korrespondierenden Gegenstrang von SEQ ID NO:33 bzw. vom kodierenden Strang gemäß SEQ ID NO:35 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem Protein mit einer ATPase oder ATPase-ähnlichen Funktion aus *S. cerevisiae*.

- 20 i) eine, vorzugsweise hochregulierte Nukleinsäuresequenz, welche für ein Protein mit der Funktion vergleichbar zu derjenigen eines PHO85-Proteins aus *S. cerevisiae* kodiert. PHO85 ist eine Kinase und ist an verschiedenen zellulären Prozessen, einschließlich der Regulation des PHO-Gens, dem Glykogen-Metabolismus, der Regulation des Zell-Zyklus und der Zell-Morphologie beteiligt.

25

Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 56“ trägt.

- 30 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 56v“ trägt.

- Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine
35 Nukleinsäuresequenz gemäß SEQ ID NO: 37. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 40 oder ein Fragment davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*,

insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

- 5 Die Inserts von „Oligo 56“ und „Oligo 56v“ besitzen mit dem MIPS Tag „Ypl110c“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 37 bzw. SEQ ID NO: 40. Die vom korrespondierenden Gegenstrang zu SEQ ID NO: 37 bzw. vom kodierenden Strang gemäß SEQ ID NO: 40 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem PHO85-Protein aus *S. cerevisiae*.

- 10 k) eine, vorzugsweise hochregulierte Nukleinsäuresequenz, welche für ein Protein mit der Funktion vergleichbar zu der eines am Membranverkehr beteiligten p24-Proteins aus *S. cerevisiae* kodiert. Mitglieder der p24 Proteinfamilie sind kleine Typ I- Transmembranproteine mit kurzem cytoplasmatischem COOH-Terminus. Diese üben eine Transportfunktion im frühen sekretorischen Weg aus und sind z.B. am Transport verschiedener sekretorischer Proteine aus dem endoplasmatischen Retikulum zum Golgi-Apparat beteiligt.

- 20 Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde ein DNA-Klon isoliert, welche für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 167“ trägt.

- 25 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 167v“ trägt.

- 30 Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 42. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 44 oder ein Fragment davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

- 35 Die Inserts von „Oligo 167“ und „Oligo 167v“ besitzen mit dem MIPS Tag „ERP5“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 42 bzw. SEQ ID NO: 44. Die vom korrespondierenden Gegenstrang zur SEQ ID NO: 42

bzw. vom kodierenden Strang der SEQ ID NO: 44 abgeleiteten Aminosäuresequenzen besitzen signifikante Sequenzhomologie mit dem am Membranverkehr beteiligten p24-Protein aus *S. cerevisiae*.

- 5 Ein weiterer Gegenstand der Erfindung betrifft Oligonukleotide, welche mit einem der obigen Polynukleotide, insbesondere unter stringenten Bedingungen, hybridisieren.

Gegenstand der Erfindung sind weiterhin Polynukleotide, welche mit einem der erfindungsgemäßen Oligonukleotide hybridisieren und für ein Genprodukt aus Mikroorganismen der Gattung
10 *Ashbya* oder ein funktionales Äquivalent dieses Genproduktes kodieren.

Die Erfindung betrifft weiterhin Polypeptide bzw. Proteine, welche von den oben beschriebenen Polynukleotiden kodiert werden; sowie Peptidfragmente davon, welche eine Aminosäuresequenz aufweisen, die wenigstens 10 zusammenhängende Aminosäurereste gemäß SEQ ID NO: 2, 4,
15 6, 7, 9, 11, 13, 15, 16, 18, 20, 22, 24, 27, 29, 30, 32, 34, 36, 38, 39, 41, 43 oder SEQ ID NO: 45 umfasst; sowie funktionale Äquivalente der erfindungsgemäßen Polypeptide bzw. Proteine.

Funktionale Äquivalente unterscheiden sich dabei von den erfindungsgemäß konkret offenbarten Produkten in ihrer Aminosäuresequenz durch Addition, Insertion, Substitution, Deletion oder
20 Inversion an wenigstens einer, wie z.B. 1 bis 30 oder 1 bis 20 oder 1 bis 10, Sequenzpositionen ohne die ursprünglich beobachtete und durch Sequenzvergleich mit anderen Proteinen ableitbare Proteinfunktion zu verlieren. Damit können Äquivalente im wesentlichen identische, höhere oder niedrigere Aktivitäten im Vergleich zum nativen Protein besitzen.

25 Weitere Gegenstände der Erfindung betreffen Expressionskassetten zur rekombinanten Produktion erfindungsgemäßer Proteine, umfassend in operativer Verknüpfung mit wenigstens einer regulativen Nukleinsäuresequenz eine der oben definierten Nukleinsäuresequenzen; sowie rekombinante Vektoren, umfassend wenigstens eine solche erfindungsgemäße Expressionskassette.

30

Erfindungsgemäß bereitgestellt werden außerdem prokaryotische oder eukaryotische Wirte, welche mit wenigstens einem Vektor obigen Typs transformiert sind. Gemäß einer bevorzugten Ausführungsform werden solche prokaryotischen oder eukaryotischen Wirte bereitgestellt, in
35 welchen die funktionale Expression wenigstens eines Gens moduliert (z.B. Inhibition oder Überexpression) ist, das für ein erfindungsgemäßes Polypeptid nach obiger Definition kodiert; oder in welchen die biologische Aktivität eines Polypeptids nach obiger Definition erniedrigt oder erhöht

ist. Bevorzugte Wirte sind ausgewählt unter Ascomyceten (Schlauchpilzen), insbesondere solchen der Gattung *Ashbya* und vorzugsweise Stämmen von *A. gossypii*.

5 Eine Modulation der Genexpression im obigen Sinn umfasst sowohl deren Inhibition, z.B. durch Blockade einer Expressionsstufe (insbesondere Transkription oder Translation) oder eine gezielte Überexpression eines Gens (z.B. durch Modifikation regulativer Sequenzen oder Erhöhung der Kopienzahl der kodierenden Sequenz).

10 Ein weiterer Gegenstand der Erfindung betrifft die Verwendung einer erfindungsgemäßen Expressionskassette, eines erfindungsgemäßen Vektors oder eines erfindungsgemäßen Wirts zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon.

15 Ein weiterer Gegenstand der Erfindung betrifft die Verwendung einer erfindungsgemäßen Expressionskassette, eines erfindungsgemäßen Vektors oder eines erfindungsgemäßen Wirts zur rekombinanten Herstellung eines erfindungsgemäßen Polypeptids nach obiger Definition.

20 Erfindungsgemäß wird weiterhin ein Verfahren zur Nachweis bzw. zur Validierung eines Effektor-targets für die Modulation der mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon bereitgestellt. Dabei behandelt man einen Mikroorganismus, der zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon befähigt ist, mit einem Effektor, welcher mit einem Target, ausgewählt unter einem erfindungsgemäßen Polypeptid nach obiger Definition oder einer dafür kodierenden Nukleinsäuresequenz, wechselwirkt (wie z.B. an diese nicht-kovalent bindet), den Einfluß des Effektors auf die Menge des mikrobiologisch produzierten Vitamins B2 und/oder des Präkursors und/oder eines
25 Derivats davon validiert; und das Target gegebenenfalls isoliert. Die Validierung erfolgt dabei bevorzugt durch direkten Vergleich mit der mikrobiologischen Vitamin B2-Produktion in Abwesenheit des Effektors unter ansonsten gleichen Bedingungen.

30 Ein weiterer Gegenstand der Erfindung betrifft ein Verfahren zur Modulation (in Bezug auf Menge und/oder Geschwindigkeit) der mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon, wobei man einen Mikroorganismus, der zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon befähigt ist, mit einem Effektor behandelt, welcher mit einem Target, ausgewählt unter einem erfindungsgemäßen Polypeptid nach obiger Definition oder einer dafür kodierenden Nukleinsäuresequenz, wechselwirkt.
35

Als bevorzugte Beispiele für oben genannte Effektoren sind zu nennen:

- a) Antikörper oder antigenbindende Fragmenten davon;
- b) von a) verschiedenen Polypeptid-Liganden, welche mit einem erfindungsgemäßen Polypeptid wechselwirken;
- c) niedermolekulare Effektoren, welche die biologische Aktivität eines erfindungsgemäßen Polypeptids modulieren;
- d) Antisense-Nukleinsäuresequenzen, welche mit einer erfindungsgemäßen Nukleinsäuresequenz wechselwirken.

Oben genannte Effektoren mit Spezifität für wenigstens eines der oben definierten erfindungsgemäßen Tragets sind ebenfalls Gegenstand der Erfindung.

Ein weiterer Gegenstand der Erfindung betrifft ein Verfahren zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon, wobei man einen Wirt gemäß obiger Definition unter die Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon begünstigenden Bedingungen kultiviert und das(die) gewünschte(n) Produkt(e) aus dem Kulturansatz isoliert. Bevorzugt ist dabei, dass man den Wirt vor und/oder während der Kultivierung mit einem Effektor nach obiger Definition behandelt. Ein bevorzugter Wirt ist dabei ausgewählt unter Mikroorganismen der Gattung *Ashbya*; insbesondere transformiert, wie oben beschrieben.

Ein letzter Gegenstand der Erfindung betrifft die Verwendung eines erfindungsgemäßen Polynukleotids oder Polypeptids als Target zur Modulation der Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon in einem Mikroorganismus der Gattung *Ashbya*.

Figurenbeschreibung:

Figur 1 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 609 bis 1 in SEQ ID NO:1) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Ygr257c aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 2 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (SEQ ID NO: 6) (entsprechend dem Gegenstrang zu Position 1494 bis 1387 in SEQ ID NO:5) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Mdl2 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 3 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem kodierenden Strang in Position 15 bis 455 in SEQ ID NO:10) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Prp12 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 4 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem kodierenden Strang in Position 246 bis 1118 in SEQ ID NO:14) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Flx1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 5 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem kodierenden Strang in Position 2 bis 790 in SEQ ID NO:19) (obere Sequenz) und einer Teilsequenz des MIPS-Tags STV1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 6 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 869 bis 522 in SEQ ID NO:23) (obere Sequenz) und einer Teilsequenz des MIPS-Tags OPT2 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 7A zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 356 bis 243 in SEQ ID NO:28) (obere Sequenz) und einer Teilsequenz des MIPS-Tags VAC1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet. Figur 7B zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 166 bis 2 in SEQ ID NO:28) (obere Sequenz) und einer Teilsequenz des MIPS-Tags VAC1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 8 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 904 bis 707 in SEQ ID NO:33) (obere Sequenz) und

einer Teilsequenz des MIPS-Tags Ymr162c aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

- 5 Figur 9 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 898 bis 5 in SEQ ID NO:37) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Ypl110c aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

10

Figur 10 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 931 bis 806 in SEQ ID NO:42) (obere Sequenz) und einer Teilsequenz des MIPS-Tags ERP5 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

15

Detaillierte Beschreibung der Erfindung:

- Die erfindungsgemäßen Nukleinsäuremoleküle kodieren Proteine bzw. Proteine, die hier als
- 20 Proteine des Transmembran-Transports (z.B. mit Aktivität bezüglich der transmembranen Transportsysteme) oder kurz als „TMT-Proteine“ bezeichnet werden. Diese TMT-Proteine haben z.B. eine Funktion bei der Steuerung membranständiger Transportsysteme, die gewünschte Proteine unter Energieverbrauch gegen einen Konzentrationsgradienten in die Zelle transportieren. Die TMT-Proteine können die Zellantwort auf äußere Bedingungen beeinflussen und dadurch
- 25 z.B. den Stoffwechsel der Zelle regulieren. Aufgrund der Verfügbarkeit von in *Ashbya gossypii* verwendbaren Klonierungsvektoren, wie z.B. offenbart in Wright und Philipsen (1991) Gene, 109, 99–105., und von Techniken zur genetischen Manipulation von *A. gossypii* und den verwandten Hefe-Arten lassen sich die erfindungsgemäßen Nukleinsäuremoleküle zur genetischen Manipulation dieser Organismen, insbesondere von *A. gossypii* verwenden, um sie als Produ-
- 30 zenten von Vitamin B2 und/oder Präkursoren und/oder Derivate davon besser und effizienter zu machen. Diese verbesserte Produktion oder Effizienz kann aufgrund einer direkten Wirkung der Manipulation eines erfindungsgemäßen Gens oder aufgrund einer indirekten Wirkung einer solchen Manipulation erfolgen.

- 35 Die vorliegende Erfindung beruht auf der Bereitstellung neuer Molekülen, die hier als TMT-Nukleinsäuren und TMT-Proteine bezeichnet werden und an dem Transmembran-Transport, insbesondere in *Ashbya gossypii*, (z.B. bei der Synthese oder Regulation von Transportprotei-

nen) beteiligt sind. Die Aktivität der erfindungsgemäßen TMT-Moleküle in *A. gossypii* beeinflusst die Vitamin B2-Produktion durch diesen Organismus. Vorzugsweise wird die Aktivität der erfindungsgemäßen TMT-Moleküle so moduliert, dass die Stoffwechsel- und/oder Energiewege von *A. gossypii*, an denen die erfindungsgemäßen TMT-Proteine teilnehmen, hinsichtlich der Ausbeute, Produktion und/oder Effizienz der Vitamin B2-Produktion moduliert werden, was entweder direkt oder indirekt die Ausbeute, Produktion und/oder Effizienz der Vitamin B2-Produktion in *A. gossypii* moduliert.

Die erfindungsgemäß bereitgestellten Nukleinsäuresequenzen sind beispielsweise aus dem Genom eines *Ashbya gossypii*-Stammes isolierbar, der von der American Type Culture Collection unter der Bezeichnung ATCC 10895 frei erhältlich ist.

Verbesserung der Vitamin B2-Produktion:

Es gibt eine Reihe von möglichen Mechanismen, über welche man durch Veränderung von Menge und/oder Aktivität eines erfindungsgemäßen MTM-Proteins die Ausbeute, Produktion und/oder Effizienz der Produktion von Vitamin B2 durch einem *A. gossypii*-Stamm direkt beeinflussen kann.

So kann durch ein effizienterer Transmembran-Transport, die Zellantwort der Zelle verstärkt und damit die Bildung von gewünschten Wertprodukten erhöht werden.

Die Mutagenese von einem oder mehreren erfindungsgemäßen TMT-Proteinen kann auch zu TMT-Proteinen mit geänderten (erhöhten oder verminderten) Aktivitäten führen, die indirekt die Produktion des gewünschten Produkts aus *A. gossypii* beeinflussen. Beispielsweise kann man mit Hilfe der TMT-Proteine die Zellen an neue oder veränderte äußere Bedingungen anpassen. Durch Verbesserung von Wachstums und Vermehrung dieser veränderter Zellen ist es möglich, die Lebensfähigkeit der Zellen in Kulturen im Großmaßstab zu steigern und auch die Teilungsrate zu verbessern. Schließlich kann dadurch die Ausbeute der gewünschten Zielprodukte, die von diesen Zellen hergestellt werden, erhöht werden.

Polypeptide:

Gegenstand der Erfindung sind Polypeptide, welche die oben genannten Aminosäuresequenzen oder charakteristische Teilsequenzen davon umfassen und/oder von den hierin beschriebenen Nukleinsäuresequenzen kodiert werden.

Erfindungsgemäß mit umfasst sind ebenfalls „funktionale Äquivalente“ der konkret offenbarten neuen Polypeptide.

- 5 „Funktionale Äquivalente“ oder Analoga der konkret offenbarten Polypeptide sind im Rahmen der vorliegenden Erfindung davon verschiedene Polypeptide, welche weiterhin die gewünschte biologische Aktivität, (wie z.B. Substratspezifität) besitzen.

- 10 Unter „funktionalen Äquivalenten“ versteht man erfindungsgemäß insbesondere Mutanten, welche in wenigstens einer der oben genannten Sequenzpositionen eine andere als die konkret genannte Aminosäure aufweisen aber trotzdem eine der oben genannten biologische Aktivitäten besitzen. „Funktionale Äquivalente“ umfassen somit die durch eine oder mehrere Aminosäure-Additionen, -Substitutionen, -Deletionen und/oder -Inversionen erhältlichen Mutanten, wobei die genannten Veränderungen in jeglicher Sequenzposition auftreten können, solange sie zu einer Mutante mit dem erfindungsgemäßen Eigenschaftsprofil führen. Funktionale Äquivalenz ist ins-
15 besondere auch dann gegeben, wenn die Reaktivitätsmuster zwischen Mutante und unverändertem Polypeptid qualitativ übereinstimmen, d.h. beispielsweise gleiche Substrate mit unterschiedlicher Geschwindigkeit umgesetzt werden.

- 20 „Funktionale Äquivalente“ im obigen Sinne sind auch Präkursoren der beschriebenen Polypeptide sowie funktionale Derivate und Salze der Polypeptide. Unter dem Ausdruck „Salze“ versteht man sowohl Salze von Carboxylgruppen als auch Säureadditionssalze von Aminogruppen der erfindungsgemäßen Proteinmoleküle. Salze von Carboxylgruppen können in an sich bekannter Weise hergestellt werden und umfassen anorganische Salze, wie zum Beispiel Natrium-, Calcium-, Ammonium-, Eisen- und Zinksalze, sowie Salze mit organischen
25 Basen, wie zum Beispiel Aminen, wie Triethanolamin, Arginin, Lysin, Piperidin und dergleichen. Säureadditionssalze, wie zum Beispiel Salze mit Mineralsäuren, wie Salzsäure oder Schwefelsäure und Salze mit organischen Säuren, wie Essigsäure und Oxalsäure sind ebenfalls Gegenstand der Erfindung.

- 30 „Funktionale Derivate“ erfindungsgemäßer Polypeptide können an funktionellen Aminosäure-Seitengruppen oder an deren N- oder C-terminalen Ende mit Hilfe bekannter Techniken ebenfalls hergestellt werden. Derartige Derivate umfassen beispielsweise aliphatische Ester von Carbonsäuregruppen, Amide von Carbonsäuregruppen, erhältlich durch Umsetzung mit Ammoniak oder mit einem primären oder sekundären Amin; N-Acylderivate freier Aminogruppen, hergestellt
35 durch Umsetzung mit Acylgruppen; oder O-Acylderivate freier Hydroxylgruppen, hergestellt durch Umsetzung mit Acylgruppen.

"Funktionale Äquivalente" umfassen natürlich auch Polypeptide welche aus anderen Organismen, zugänglich sind, sowie natürlich vorkommende Varianten. Beispielsweise lassen sich durch Sequenzvergleich Bereiche homologer Sequenzregionen festlegen und in Anlehnung an die konkreten Vorgaben der Erfindung äquivalente Enzyme ermitteln.

5

„Funktionale Äquivalente“ umfassen ebenfalls Fragmente, vorzugsweise einzelne Domänen oder Sequenzmotive, der erfindungsgemäßen Polypeptide, welche z.B. die gewünschte biologische Funktion aufweisen.

- 10 „Funktionale Äquivalente“ sind außerdem Fusionsproteine, welche ein der oben genannten Polypeptidsequenzen oder davon abgeleitete funktionale Äquivalente und wenigstens eine weitere, davon funktionell verschiedene, heterologe Sequenz in funktioneller N- oder C-terminaler Verknüpfung (d.h. ohne gegenseitigen wesentliche funktionelle Beeinträchtigung der Fusionsproteinteile) aufweisen. Nichtlimitierende Beispiele für derartige heterologe Sequenzen sind z.B. Signalpeptide, Enzyme, Immunoglobuline, Oberflächenantigene, Rezeptoren oder Rezeptorliganden.
- 15

- Erfindungsgemäß mit umfasste „funktionale Äquivalente“ sind Homologe zu den konkret offenbarten Proteinen. Diese besitzen wenigstens 60 %, vorzugsweise wenigstens 75% ins besondere wenigstens 85 %, wie z.B. 90%, 95% oder 99%, Homologie zu einer der konkret offenbarten Sequenzen, berechnet nach dem Algorithmus von Pearson und Lipman, Proc. Natl. Acad. Sci. (USA) 85(8), 1988, 2444-2448.
- 20

- Im Falle einer möglichen Proteinglykosylierung umfassen erfindungsgemäße Äquivalente Proteine des oben bezeichneten Typs in deglykosylierter bzw. glykosylierter Form sowie durch Veränderung des Glykosylierungsmusters erhältliche abgewandelte Formen.
- 25

- Homologe der erfindungsgemäßen Proteine oder Polypeptide können durch Mutagenese erzeugt werden, z.B. durch Punktmutation oder Verkürzung des Proteins. Der Begriff "Homolog", wie er hier verwendet wird, betrifft eine variante Form des Proteins, die als Agonist oder Antagonist der Protein-Aktivität wirkt.
- 30

- Homologe der erfindungsgemäßen Proteine können durch Screening kombinatorischer Banken von Mutanten, wie z.B. Verkürzungsmutanten, identifiziert werden. Beispielsweise kann eine variierte Bank von Protein-Varianten durch kombinatorische Mutagenese auf Nukleinsäureebene erzeugt werden, wie z.B. durch enzymatisches Ligieren eines Gemisches synthetischer Oligonukleotide. Es gibt eine Vielzahl von Verfahren, die zur Herstellung von Banken potentieller
- 35

Homologer aus einer degenerierten Oligonukleotidsequenz verwendet werden können. Die chemische Synthese einer degenerierten Gensequenz kann in einem DNA-Syntheseautomaten durchgeführt werden, und das synthetische Gen kann dann in einen geeigneten Expressionsvektor ligiert werden. Die Verwendung eines degenerierten Gensatzes ermöglicht die Bereitstellung sämtlicher Sequenzen in einem Gemisch, die den gewünschten Satz an potentiellen Proteinsequenzen codieren. Verfahren zur Synthese degenerierter Oligonukleotide sind dem Fachmann bekannt (Z.B. Narang, S.A. (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al., (1984) Science 198:1056; Ike et al. (1983) Nucleic Acids Res. 11:477).

Zusätzlich können Banken von Fragmenten des Protein-Codons verwendet werden, um eine variierte Population von Protein-Fragmenten zum Screening und zur anschließenden Selektion von Homologen eines erfindungsgemäßen Proteins zu erzeugen. Bei einer Ausführungsform kann eine Bank von kodierenden Sequenzfragmenten durch Behandeln eines doppelsträngigen PCR-Fragmentes einer kodierenden Sequenz mit einer Nuklease unter Bedingungen, unter denen ein Nicking nur etwa einmal pro Molekül erfolgt, Denaturieren der doppelsträngigen DNA, Renaturieren der DNA unter Bildung doppelsträngiger DNA, die Sense-/Antisense-Paare von verschiedenen genickten Produkten umfassen kann, Entfernen einzelsträngiger Abschnitte aus neu gebildeten Duplices durch Behandlung mit S1-Nuclease und Ligieren der resultierenden Fragmentbank in einen Expressionsvektor erzeugt werden. Durch dieses Verfahren kann eine Expressionsbank hergeleitet werden, die N-terminale, C-terminale und interne Fragmente mit verschiedenen Größen des erfindungsgemäßen Proteins kodiert.

Im Stand der Technik sind mehrere Techniken zum Screening von Genprodukten kombinatorischer Banken, die durch Punktmutationen oder Verkürzung hergestellt worden sind, und zum Screening von cDNA-Banken auf Genprodukte mit einer ausgewählten Eigenschaft bekannt. Diese Techniken lassen sich an das schnelle Screening der Genbanken anpassen, die durch kombinatorische Mutagenese von erfindungsgemäßer Homologen erzeugt worden sind. Die am häufigsten verwendeten Techniken zum Screening großer Genbanken, die einer Analyse mit hohem Durchsatz unterliegen, umfassen das Klonieren der Genbank in replizierbare Expressionsvektoren, Transformieren der geeigneten Zellen mit der resultierenden Vektorenbank und Exprimieren der kombinatorischen Gene unter Bedingungen, unter denen der Nachweis der gewünschten Aktivität die Isolation des Vektors, der das Gen codiert, dessen Produkt nachgewiesen wurde, erleichtert. Recursive-Ensemble-Mutagenese (REM), eine Technik, die die Häufigkeit funktioneller Mutanten in den Banken vergrößert, kann in Kombination mit den Screeningtests verwendet werden, um Homologe zu identifizieren (Arkin und Yourvan (1992) PNAS 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3):327-331).

Die erfindungsgemäßen Polypeptide können rekombinant hergestellt werden (vgl. folgende Abschnitte) oder können in nativer Form unter Anwendung klassischer biochemischer Arbeitsweisen (vgl. Cooper, T. G., Biochemische Arbeitsmethoden, Verlag Walter de Gruyter, Berlin, New York oder in Scopes, R., Protein Purification, Springer Verlag, New York, Heidelberg, Berlin) aus Mikroorganismen, insbesondere solchen der Gattung Ashbya, isoliert werden.

Nukleinsäuresequenzen:

10 Gegenstand der Erfindung sind auch Nukleinsäuresequenzen (einzeln- und doppelsträngige DNA- und RNA-Sequenzen, wie z.B. cDNA und mRNA), kodierend für eines der obigen Polypeptide und deren funktionalen Äquivalenten, welche z.B. unter Verwendung künstlicher Nukleotidanaloga zugänglich sind.

15 Die Erfindung betrifft sowohl isolierte Nukleinsäuremoleküle, welche für erfindungsgemäße Polypeptide bzw. Proteine oder biologisch aktive Abschnitte davon kodieren, sowie Nukleinsäurefragmente, die z.B. zur Verwendung als Hybridisierungs-sonden oder Primer zur Identifizierung oder Amplifizierung von erfindungsgemäßen kodierenden Nukleinsäuren verwendet werden können.

20 Die erfindungsgemäßen Nukleinsäuremoleküle können zudem untranslatierte Sequenzen vom 3'- und/oder 5'-Ende des kodierenden Genbereichs enthalten.

25 Ein "isoliertes" Nukleinsäuremolekül wird von anderen Nukleinsäuremolekülen abgetrennt, die in der natürlichen Quelle der Nukleinsäure zugegen sind und kann überdies im wesentlichen frei von anderem zellulären Material oder Kulturmedium sein, wenn es durch rekombinante Techniken hergestellt wird, oder frei von chemischen Vorstufen oder anderen Chemikalien sein, wenn es chemisch synthetisiert wird.

30 Ein erfindungsgemäßes Nukleinsäuremolekül kann mittels molekularbiologischer Standard-Techniken und der erfindungsgemäß bereitgestellten Sequenzinformation isoliert werden. Beispielsweise kann cDNA aus einer geeigneten cDNA-Bank isoliert werden, indem eine der konkret offenbarten vollständigen Sequenzen oder ein Abschnitt davon als Hybridisierungs-sonde und Standard-Hybridisierungstechniken (wie z.B. beschrieben in Sambrook, J., Fritsch, E.F. und
35 Maniatis, T. Molecular Cloning: A Laboratory Manual. 2. Aufl., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) verwendet werden. Überdies läßt sich ein Nukleinsäuremolekül, umfassend eine der offenbarten Sequenzen oder einen

Abschnitt davon, durch Polymerasekettenreaktion isolieren, wobei die Oligonukleotidprimer, die auf der Basis dieser Sequenz erstellt wurden, verwendet werden. Die so amplifizierte Nukleinsäure kann in einen geeigneten Vektor kloniert werden und durch DNA-Sequenzanalyse charakterisiert werden. Die erfindungsgemäßen Oligonukleotide, die einer TMT-Nukleotidsequenz entsprechen, können ferner durch Standard-Syntheseverfahren, z.B. mit einem automatischen DNA-Synthesegerät, hergestellt werden.

Die Erfindung umfasst weiterhin die zu den konkret beschriebenen Nukleotidsequenzen komplementären Nukleinsäuremoleküle oder einen Abschnitt davon.

Die erfindungsgemäßen Nukleotidsequenzen ermöglichen die Erzeugung von Sonden und Primern, die zur Identifizierung und/oder Klonierung homologer Sequenzen in anderen Zelltypen und Organismen verwendbar sind. Solche Sonden bzw. Primer umfassen gewöhnlich einen Nukleotidsequenzbereich, der unter stringenten Bedingungen an mindestens etwa 12, vorzugsweise mindestens etwa 25, wie z.B. etwa 40, 50 oder 75 aufeinanderfolgende Nukleotide eines Sense-Stranges einer erfindungsgemäßen Nukleinsäuresequenz oder eines entsprechenden Antisense-Stranges hybridisiert.

Weitere erfindungsgemäße Nukleinsäuresequenzen sind abgeleitet von SEQ ID NO: 1, 3, 5, 8, 10, 12, 14, 17, 19, 21, 23, 25, 26, 28, 31, 33, 35, 37, 40, 42 oder SEQ ID NO: 44 und unterscheiden sich davon durch Addition, Substitution, Insertion oder Deletion einzelner oder mehrerer Nukleotide, kodieren aber weiterhin für Polypeptide mit dem gewünschten Eigenschaftsprofil.

Erfindungsgemäß umfasst sind auch solche Nukleinsäuresequenzen, die sogenannte stumme Mutationen umfassen oder entsprechend der Codon-Nutzung eines speziellen Ursprungs- oder Wirtsorganismus, im Vergleich zu einer konkret genannten Sequenz verändert sind, ebenso wie natürlich vorkommende Varianten, wie z.B. Spleißvarianten oder Allelvarianten, davon. Gegenstand sind ebenso durch konservative Nukleotidsubstitutionen (d.h. die betreffende Aminosäure wird durch eine Aminosäure gleicher Ladung, Größe, Polarität und/oder Löslichkeit ersetzt) erhältliche Sequenzen.

Gegenstand der Erfindung sind auch die durch Sequenzpolymorphismen von den konkret offenbarten Nukleinsäuren abgeleiteten Moleküle. Diese genetischen Polymorphismen können zwischen Individuen innerhalb einer Population aufgrund der natürlichen Variation existieren. Diese natürlichen Variationen bewirken üblicherweise eine Varianz von 1 bis 5 % in der Nukleotidsequenz eines Gens.

Weiterhin umfasst die Erfindung auch Nukleinsäuresequenzen, welchen mit oben genannten kodierenden Sequenzen hybridisieren oder dazu komplementär sind. Diese Polynukleotide lassen sich bei Durchmusterung von genomischen oder cDNA-Banken auffinden und gegebenenfalls daraus mit geeigneten Primern mittels PCR vermehren und anschließend beispielsweise mit geeigneten Sonden isolieren. Eine weitere Möglichkeit bietet die Transformation geeigneter Mikroorganismen mit erfindungsgemäßen Polynukleotiden oder Vektoren, die Vermehrung der Mikroorganismen und damit der Polynukleotide und deren anschließende Isolierung. Darüber hinaus können erfindungsgemäße Polynukleotide auch auf chemischem Wege synthetisiert werden.

Unter der Eigenschaft, an Polynukleotide „hybridisieren“ zu können, versteht man die Fähigkeit eines Poly- oder Oligonukleotids unter stringenten Bedingungen an eine nahezu komplementäre Sequenz zu binden, während unter diesen Bedingungen unspezifische Bindungen zwischen nicht-komplementären Partnern unterbleiben. Dazu sollten die Sequenzen zu 70-100%, vorzugsweise zu 90-100%, komplementär sein. Die Eigenschaft komplementärer Sequenzen, spezifisch aneinander binden zu können, macht man sich beispielsweise in der Northern- oder Southern-Blot-Technik oder bei der Primerbindung in PCR oder RT-PCR zunutze. Üblicherweise werden dazu Oligonukleotide ab einer Länge von 30 Basenpaaren eingesetzt. Unter stringenten Bedingungen versteht man beispielsweise in der Northern-Blot-Technik die Verwendung einer 50 – 70 °C, vorzugsweise 60 – 65 °C warmen Waschlösung, beispielsweise 0,1x SSC-Puffer mit 0,1% SDS (20x SSC: 3M NaCl, 0,3M Na-Citrat, pH 7,0) zur Elution unspezifisch hybridisierter cDNA-Sonden oder Oligonukleotide. Dabei bleiben, wie oben erwähnt, nur in hohem Maße komplementäre Nukleinsäuren aneinander gebunden. Die Einstellung stringenter Bedingungen ist dem Fachmann bekannt und ist z.B. in Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. beschrieben.

Ein weiterer Aspekt der Erfindung betrifft „Antisense“-Nukleinsäuren. Diese umfaßt eine Nukleotidsequenz, die zu einer kodierenden „Sense“-Nukleinsäure, komplementär ist. Die Antisense-Nukleinsäure kann zum gesamten kodierenden Strang oder nur zu einem Abschnitt davon komplementär sein. Bei einer weiteren Ausführungsform ist das Antisense-Nukleinsäuremolekül antisense zu einem nicht-kodierenden Bereich des kodierenden Stranges einer Nukleotidsequenz. Der Begriff „nicht-kodierender Bereich“ betrifft die als 5'- und 3'-untranslatierte Bereiche bezeichneten Sequenzabschnitte.

Ein Antisense-Oligonukleotid kann bspw. etwa 5, 10, 15, 20, 25, 30, 35, 40, 45 oder 50 Nukleotide lang sein. Eine erfindungsgemäße Antisense-Nukleinsäure kann durch chemische Synthese

und enzymatische Ligationsreaktionen mittels im Fachgebiet bekannter Verfahren konstruiert werden. Eine Antisense-Nukleinsäure kann chemisch synthetisiert werden, wobei natürlich vorkommende Nukleotide oder verschieden modifizierte Nukleotide verwendet werden, die so gestaltet sind, dass sie die biologische Stabilität der Moleküle erhöhen, oder die physikalische Stabilität des Duplexes erhöhen, der zwischen der Antisense- und Sense-Nukleinsäure entstanden ist. Beispielsweise können Phosphorthioat-Derivate und acridinsubstituierte Nukleotide verwendet werden. Beispiele modifizierter Nukleoside, die zur Erzeugung der Antisense-Nukleinsäure verwendet werden können, sind u.a. 5-Fluoruracil, 5-Bromuracil, 5-Chloruracil, 5-Ioduracil, Hypoxanthin, Xanthin, 4-Acetylcytosin, 5-(Carboxyhydroxymethyl)uracil, 5-Carboxymethylaminomethyl-2-thiouridin, 5-Carboxymethylaminomethyluracil, Dihydrouracil, Beta-D-Galactosylqueosin, Inosin, N6-Isopentenyladenin, 1-Methylguanin, 1-Methylinosin, 2,2-Dimethylguanin, 2-Methyladenin, 2-Methylguanin, 3-Methylcytosin, 5-Methylcytosin, N6-Adenin, 7-Methylguanin, 5-Methylaminomethyluracil, 5-Methoxyaminomethyl-2-thiouracil, Beta-D-Mannosylqueosin, 5'-Methoxycarboxymethyluracil, 5-Methoxyuracil, 2-Methylthio-N6-isopentenyladenin, Uracil-5-oxyessigsäure (v), Wybutoxosin, Pseudouracil, Queosin, 2-Thiocytosin, 5-Methyl-2-thiouracil, 2-Thiouracil, 4-Thiouracil, 5-Methyluracil, Uracil-5-oxyessigsäuremethylester, 3-(3-Amino-3-N-2-carboxypropyl)uracil, (acp3)w und 2,6-Diaminopurin. Die Antisense-Nukleinsäure kann auch biologisch hergestellt werden, indem ein Expressionsvektor verwendet wird, in den eine Nukleinsäure in Antisense-Richtung subkloniert worden ist.

Die erfindungsgemäßen Antisense-Nukleinsäuremoleküle werden üblicherweise an eine Zelle verabreicht oder in situ erzeugt, so dass sie mit der zellulären mRNA und/oder einer kodierenden DNA hybridisieren oder daran binden, so dass die Expression des Proteins, z.B. durch Hemmung der Transkription und/oder Translation, gehemmt wird.

Das Antisense-Molekül kann so modifiziert werden, dass es spezifisch an einen Rezeptor oder an ein Antigen bindet, das auf einer ausgewählten Zelloberfläche exprimiert wird, z.B. durch Verknüpfen des Antisense-Nukleinsäuremoleküls mit einem Peptid oder einem Antikörper, das/der an einen Zelloberflächenrezeptor oder Antigen bindet. Das Antisense-Nukleinsäuremolekül kann auch unter Verwendung der hier beschriebenen Vektoren an Zellen verabreicht werden. Zur Erzielung hinreichender intrazellulärer Konzentrationen der Antisense-Moleküle sind Vektorkonstrukte, in denen sich das Antisense-Nukleinsäuremolekül unter der Kontrolle eines starken bakteriellen, viralen oder eukaryotischen Promotors befindet, bevorzugt.

In einer weiteren Ausführungsform ist das erfindungsgemäße Antisense-Nukleinsäuremolekül ein alpha-anomeres Nukleinsäuremolekül. Ein alpha-anomeres Nukleinsäuremolekül bildet spe-

zifische doppelsträngige Hybride mit komplementärer RNA, wobei die Stränge im Gegensatz zu gewöhnlichen alpha-Einheiten parallel zueinander verlaufen. (Gaultier et al., (1987) Nucleic Acids Res. 15:6625-6641). Das Antisense-Nukleinsäuremolekül kann zudem ein 2'-O-Methylribonukleotid (Inoue et al., (1987) Nucleic Acids Res. 15:6131-6148) oder ein chimäres RNA-DNA-Analogon (Inoue et al. (1987) FEBS Lett. 215:327-330) umfassen.

Gegenstand der Erfindung sind auch Ribozyme. Dies sind katalytische RNA-Moleküle mit Ribonukleaseaktivität, die eine einzelsträngige Nukleinsäure, wie eine mRNA, zu der sie einen komplementären Bereich haben, spalten können. Somit können Ribozyme (z.B. Hammerhead-Ribozyme (beschrieben in Haselhoff und Gerlach (1988) Nature 334:585-591)) zur katalytischen Spaltung von erfindungsgemäßen Transkripten verwendet werden, um dadurch die Translation der entsprechenden Nukleinsäure zu hemmen. Ein Ribozym mit Spezifität für eine erfindungsgemäße kodierende Nukleinsäure kann z.B. auf der Basis einer hierin konkret offenbaren cDNA gebildet werden. Beispielsweise kann ein Derivat einer Tetrahymena-L-19-IVS-RNA konstruiert werden, wobei die Nukleotidsequenz der aktiven Stelle komplementär zur Nukleotidsequenz ist, die in einer erfindungsgemäßen kodierenden mRNA gespalten werden soll. (vgl. z.B. US-A-4 987 071 und US-A-5 116 742). Alternativ kann mRNA zur Selektion einer katalytischen RNA mit spezifischer Ribonukleaseaktivität aus einem Pool von RNA-Molekülen verwendet werden (siehe z.B. Bartel, D., und Szostak, J.W. (1993) Science 261:1411-1418).

Die Genexpression erfindungsgemäßer Sequenzen läßt sich alternativ hemmen, indem Nukleotidsequenzen, die komplementär zum regulatorischen Bereich einer erfindungsgemäßen Nukleotidsequenz sind (z.B. zu einem Promotor und/oder Enhancer einer kodierenden Sequenz) so dirigiert werden, dass Triple-Helixstrukturen gebildet werden, die die Transkription des entsprechenden Gens in Ziel-Zellen verhindern (Helene, C. (1991) Anticancer Drug Res. 6(6) 569-584; Helene, C. et al., (1992) Ann. N. Y. Acad. Sci. 660:27-36; und Maher, L.J. (1992) Bioassays 14(12):807-815).

Expressionskonstrukte und Vektoren:

Gegenstand der Erfindung sind außerdem Expressionskonstrukte, enthaltend unter der genetischen Kontrolle regulativer Nukleinsäuresequenzen eine für ein erfindungsgemäßes Polypeptid kodierende Nukleinsäuresequenz; sowie Vektoren, umfassend wenigstens eines dieser Expressionskonstrukte. Vorzugsweise umfassen solche erfindungsgemäßen Konstrukte 5'-stromaufwärts von der jeweiligen kodierenden Sequenz einen Promotor und 3'-stromabwärts eine Terminatorsequenz sowie gegebenenfalls weitere übliche regulative Elemente, und zwar jeweils operativ verknüpft mit der kodierenden Sequenz. Unter einer „operativen Verknüpfung“ versteht

man die sequentielle Anordnung von Promotor, kodierender Sequenz, Terminator und gegebenenfalls weiterer regulativer Elemente derart, dass jedes der regulativen Elemente seine Funktion bei der Expression der kodierenden Sequenz bestimmungsgemäß erfüllen kann. Beispiele für operativ verknüpfbare Sequenzen sind Targeting-Sequenzen sowie Enhancer, Polyadenylierungssignale und dergleichen. Weitere regulative Elemente umfassen selektierbare Marker, Amplifikationssignale, Replikationsursprünge und dergleichen. Geeignete regulatorische Sequenzen sind z.B. beschrieben in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990).

10 Zusätzlich zu den artifiziellen Regulationssequenzen kann die natürliche Regulationssequenz vor dem eigentlichen Strukturgen noch vorhanden sein. Durch genetische Veränderung kann diese natürliche Regulation gegebenenfalls ausgeschaltet und die Expression der Gene erhöht oder erniedrigt werden. Das Genkonstrukt kann aber auch einfacher aufgebaut sein, das heißt es werden keine zusätzlichen Regulationssignale vor das Strukturgen insertiert und der natürliche Promotor mit seiner Regulation wird nicht entfernt. Statt dessen wird die natürliche Regulationssequenz so mutiert, dass keine Regulation mehr erfolgt und die Genexpression gesteigert oder verringert wird. Die Nukleinsäuresequenzen können in einer oder mehreren Kopien im Genkonstrukt enthalten sein.

20 Beispiele für brauchbare Promotoren sind: cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, lacIq-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, λ -PR- oder im λ -PL-Promotor, die vorteilhafterweise in gram-negativen Bakterien Anwendung finden; sowie die gram-positiven Promotoren amy und SPO2, die Hefepromotoren ADC1, MF α , AC, P-60, CYC1, GAPDH oder die Pflanzenpromotoren CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, not oder der Ubiquitin- oder Phaseolin-Promotor. Besonders bevorzugt ist die Verwendung induzierbarer Promotoren, wie z.B. licht- und insbesondere temperaturinduzierbarer Promotoren, wie der P_rP_r-Promotor. Prinzipiell können alle natürlichen Promotoren mit ihren Regulationssequenzen verwendet werden. Darüber hinaus können auch synthetische Promotoren vorteilhaft verwendet werden.

30 Die genannten regulatorischen Sequenzen sollen die gezielte Expression der Nukleinsäuresequenzen ermöglichen. Dies kann beispielsweise je nach Wirtsorganismus bedeuten, dass das Gen erst nach Induktion exprimiert oder überexprimiert wird, oder dass es sofort exprimiert und/oder überexprimiert wird.

35 Die regulatorischen Sequenzen bzw. Faktoren können dabei vorzugsweise die Expression positiv beeinflussen und dadurch erhöhen oder erniedrigen. So kann eine Verstärkung der regulatorischen Elemente vorteilhafterweise auf der Transkriptionsebene erfolgen, indem starke

Transkriptionssignale wie Promotoren und/oder "Enhancer" verwendet werden. Daneben ist aber auch eine Verstärkung der Translation möglich, indem beispielsweise die Stabilität der mRNA verbessert wird.

- 5 Die Herstellung einer Expressionskassette erfolgt durch Fusion eines geeigneten Promotors mit einer geeigneten kodierenden Nukleotidsequenz sowie einem Terminator- oder Polyadenylierungssignal. Dazu verwendet man gängige Rekombinations- und Klonierungstechniken, wie sie beispielsweise in T. Maniatis, E.F. Fritsch und J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) sowie in T.J. Silhavy, 10 M.L. Berman und L.W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) und in Ausubel, F.M. et al., *Current Protocols in Molecular Biology*, Greene Publishing Assoc. and Wiley Interscience (1987) beschrieben sind.

- Das rekombinante Nukleinsäurekonstrukt bzw. Genkonstrukt wird zur Expression in einem geeigneten Wirtsorganismus vorteilhafterweise in einen wirtsspezifischen Vektor inseriert, der eine 15 optimale Expression der Gene im Wirt ermöglicht. Vektoren sind dem Fachmann wohl bekannt und können beispielsweise aus "Cloning Vectors" (Pouwels P. H. et al., Hrsg, Elsevier, Amsterdam-New York-Oxford, 1985) entnommen werden. Unter Vektoren sind außer Plasmiden auch alle anderen dem Fachmann bekannten Vektoren, wie beispielsweise Phagen, Viren, wie SV40, 20 CMV, Baculovirus und Adenovirus, Transposons, IS-Elemente, Phasmide, Cosmide, und lineare oder zirkuläre DNA zu verstehen. Diese Vektoren können autonom im Wirtsorganismus repliziert oder chromosomal repliziert werden.

Als Beispiele für geeignete Expressionsvektoren können genannt werden:

- 25 Übliche Fusionsexpressionsvektoren, wie pGEX (Pharmacia Biotech Inc; Smith, D.B. und Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) und pRIT 5 (Pharmacia, Piscataway, NJ), bei denen Glutathion-S-Transferase (GST), Maltose E-bindendes Protein bzw. Protein A an das rekombinante Zielprotein fusioniert wird.

- 30 Nicht-Fusionsprotein-Expressionsvektoren wie pTrc (Amann et al., (1988) *Gene* 69:301-315) und pET 11d (Studier et al. *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Kalifornien (1990) 60-89).

- 35 Hefe-Expressionsvektor zur Expression in der Hefe *S. cerevisiae*, wie pYepSec1 (Baldari et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan und Herskowitz (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123) sowie pYES2 (Invitrogen Corporation, San Diego, CA).

Vektoren und Verfahren zur Konstruktion von Vektoren, die sich zur Verwendung in anderen Pilzen, wie filamentösen Pilzen, eignen, umfassen diejenigen, die eingehend beschrieben sind in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy et al., Hrsg., S. 1-28, Cambridge University Press: Cambridge.

Baculovirus-Vektoren, die zur Expression von Proteinen in gezüchteten Insektenzellen (bspw. Sf9-Zellen) verfügbar sind, umfassen die pAc-Reihe (Smith et al., (1983) Mol. Cell Biol. 3:2156-2165) und die pVL-Reihe (Lucklow und Summers (1989) Virology 170:31-39).

Pflanzen-Expressionsvektoren, wie solche, die eingehend beschrieben sind in: Becker, D., Kemper, E., Schell, J. und Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", Plant Mol. Biol. 20:1195-1197; und Bevan, M.W. (1984) "Binary Agrobacterium vectors for plant transformation", Nucl. Acids Res. 12:8711-8721.

Säugetier-Expressionsvektoren, wie pCDM8 (Seed, B. (1987) Nature 329:840) und pMT2PC (Kaufman et al. (1987) EMBO J. 6:187-195).

Weitere geeignete Expressionssysteme für prokaryontische und eukaryotische Zellen sind in Kapitel 16 und 17 von Sambrook, J., Fritsch, E.F. und Maniatis, T., Molecular cloning: A Laboratory Manual, 2. Auflage, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989 beschrieben.

Rekombinante Mikroorganismen:

Mit Hilfe der erfindungsgemäßen Vektoren sind rekombinante Mikroorganismen herstellbar, welche beispielsweise mit wenigstens einem erfindungsgemäßen Vektor transformiert sind und zur Produktion der erfindungsgemäßen Polypeptide eingesetzt werden können. Vorteilhafterweise werden die oben beschriebenen erfindungsgemäßen rekombinanten Konstrukte in ein geeignetes Wirtssystem eingebracht und exprimiert. Dabei werden vorzugsweise dem Fachmann bekannte geläufige Klonierungs- und Transfektionsmethoden, wie beispielsweise Co-Präzipitation, Protoplastenfusion, Elektroporation, retrovirale Transfektion und dergleichen, verwendet, um die genannten Nukleinsäuren im jeweiligen Expressionssystem zur Expression zu bringen. Geeignete Systeme werden beispielsweise in Current Protocols in Molecular Biology, F. Ausubel et al., Hrsg., Wiley Interscience, New York 1997, oder Sambrook et al. Molecular Cloning: A Laboratory Manual. 2. Aufl., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989 beschrieben.

Erfindungsgemäß sind auch homolog rekombinierten Mikroorganismen herstellbar. Dazu wird ein Vektor hergestellt, der zumindest einen Abschnitt eines erfindungsgemäßen Gens oder einer kodierenden Sequenz enthält, worin gegebenenfalls wenigstens eine Aminosäure-Deletion, -
5 Addition oder -Substitution eingebracht worden ist, um die erfindungsgemäße Sequenz zu verändern, z.B. funktionell zu disruptieren ("Knockout"-Vektor). Die eingebrachte Sequenz kann z.B. auch ein Homologes aus einem Verwandten Mikroorganismus sein oder aus einer Säuge-
tier-, Hefe- oder Insektenquelle abgeleitet sein. Der zur homologen Rekombination verwendete Vektor kann alternativ derart ausgestaltet sein, dass das endogene Gen bei homologer Rekombination mutiert oder anderweitig verändert ist, jedoch noch das funktionelle Protein codiert (z.B.
10 kann der stromaufwärts gelegene regulatorische Bereich derart verändert sein, dass dadurch die Expression des endogenen Proteins verändert wird). Der veränderte Abschnitt des TMT-Gens ist im homologen Rekombinationsvektor. Die Konstruktion geeigneter Vektoren zur homologen Rekombination ist z.B. beschrieben in Thomas, K.R. und Capecchi, M.R. (1987) Cell 51:503.

15 Als Wirtsorganismen sind prinzipiell alle Organismen geeignet, die eine Expression der erfindungsgemäßen Nukleinsäuren, ihrer Allelvarianten, ihrer funktionellen Äquivalente oder Derivate ermöglichen. Unter Wirtsorganismen sind beispielsweise Bakterien, Pilze, Hefen, pflanzliche oder tierische Zellen zu verstehen. Bevorzugte Organismen sind Bakterien, wie solche der Gat-
20 tungen Escherichia, wie z. B. Escherichia coli, Streptomyces, Bacillus oder Pseudomonas, eukaryotische Mikroorganismen, wie Saccharomyces cerevisiae, Aspergillus, höhere eukaryotische Zellen aus Tieren oder Pflanzen, beispielsweise Sf9 oder CHO-Zellen. Bevorzugte Organismen sind aus der Gattung Ashbya, insbesondere aus A. gossypii-Stämmen ausgewählt.

25 Die Selektion erfolgreich transformierter Organismen kann durch Markergene erfolgen, die ebenfalls im Vektor oder in der Expressionskassette enthalten sind. Beispiele für solche Markergene sind Gene für Antibiotikaresistenz und für Enzyme, die eine farbgebende Reaktion katalysieren, die ein Anfärben der transformierten Zelle bewirkt. Diese können dann mittels automatischer Zellsortierung selektiert werden. Erfolgreich mit einem Vektor transformierte Mikroorganismen,
30 die ein entsprechendes Antibiotikaresistenzgen (z.B. G418 oder Hygromycin) tragen, lassen sich durch entsprechende Antibiotika-enhaltende Medien oder Nährböden selektieren. Markerproteine, die an der Zelloberfläche präsentiert werden, können zur Selektion mittels Affinitätschromatographie genutzt werden.

35 Die Kombination aus den Wirtsorganismen und den zu den Organismen passenden Vektoren, wie Plasmide, Viren oder Phagen, wie beispielsweise Plasmide mit dem RNA-Polymerase/Promoter-System, die Phagen λ oder μ oder andere temperente Phagen oder

Transposons und/oder weiteren vorteilhaften regulatorischen Sequenzen bildet ein Expressionssystem. Beispielsweise ist unter dem Begriff "Expressionssystem" die Kombination aus Säugertierzellen, wie CHO-Zellen, und Vektoren, wie pcDNA3neo-Vektor, die für Säugertierzellen geeignet sind, zu verstehen.

5

Gewünschtenfalls kann das Genprodukt auch in transgenen Organismen wie transgenen Tieren, wie insbesondere Mäusen, Schafen oder transgenen Pflanzen zur Expression gebracht werden.

Rekombinante Herstellung der Polypeptide:

10

Gegenstand der Erfindung sind weiterhin Verfahren zur rekombinanten Herstellung einer erfindungsgemäßen Polypeptide oder funktioneller, biologisch aktiver Fragmente davon, wobei man einen Polypeptide-produzierenden Mikroorganismus kultiviert, gegebenenfalls die Expression der Polypeptide induziert und diese aus der Kultur isoliert. Die Polypeptide können so auch in großtechnischem Maßstab produziert werden, falls dies erwünscht ist.

15

Der rekombinante Mikroorganismus kann nach bekannten Verfahren kultiviert und fermentiert werden. Bakterien können beispielsweise in TB- oder LB-Medium und bei einer Temperatur von 20 bis 40°C und einem pH-Wert von 6 bis 9 vermehrt werden. Im Einzelnen werden geeignete Kultivierungsbedingungen beispielsweise in T. Maniatis, E.F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) beschrieben.

20

Die Zellen werden dann, falls die Polypeptide nicht in das Kulturmedium sezerniert werden, aufgeschlossen und das Produkt nach bekannten Proteinisolierungsverfahren aus dem Lysat gewonnen. Die Zellen können wahlweise durch hochfrequenten Ultraschall, durch hohen Druck, wie z.B. in einer French-Druckzelle, durch Osmolyse, durch Einwirkung von Detergenzien, lytischen Enzymen oder organischen Lösungsmitteln, durch Homogenisatoren oder durch Kombination mehrerer der aufgeführten Verfahren aufgeschlossen werden.

25

30

Eine Aufreinigung der Polypeptide kann mit bekannten, chromatographischen Verfahren erzielt werden, wie Molekularsieb-Chromatographie (Gelfiltration), wie Q-Sepharose-Chromatographie, Ionenaustausch-Chromatographie und hydrophobe Chromatographie, sowie mit anderen üblichen Verfahren wie Ultrafiltration, Kristallisation, Aussalzen, Dialyse und nativer Gelelektrophorese. Geeignete Verfahren werden beispielsweise in Cooper, T. G., Biochemische Arbeitsmethoden, Verlag Walter de Gruyter, Berlin, New York oder in Scopes, R., Protein Purification, Springer Verlag, New York, Heidelberg, Berlin beschrieben.

35

Besonders vorteilhaft ist es, zur Isolierung des rekombinanten Proteins Vektorsysteme oder Oligonukleotide zu verwenden, die die cDNA um bestimmte Nucleotidsequenzen verlängern und damit für veränderte Polypeptide oder Fusionsproteine kodieren, die z.B. einer einfacheren Reinigung dienen. Derartige geeignete Modifikationen sind beispielsweise als Anker fungierende sogenannte "Tags", wie z.B. die als Hexa-Histidin-Anker bekannte Modifikation oder Epitope, die als Antigene von Antikörpern erkannt werden können (beschrieben zum Beispiel in Harlow, E. and Lane, D., 1988, Antibodies: A Laboratory Manual. Cold Spring Harbor (N.Y.) Press). Diese Anker können zur Anheftung der Proteine an einen festen Träger, wie z.B. einer Polymermatrix, dienen, die beispielsweise in einer Chromatographiesäule eingefüllt sein kann, oder an einer Mikrotiterplatte oder an einem sonstigen Träger verwendet werden kann.

Gleichzeitig können diese Anker auch zur Erkennung der Proteine verwendet werden. Zur Erkennung der Proteine können außerdem übliche Marker, wie Fluoreszenzfarbstoffe, Enzymmarker, die nach Reaktion mit einem Substrat ein detektierbares Reaktionsprodukt bilden, oder radioaktive Marker, allein oder in Kombination mit den Ankern zur Derivatisierung der Proteine verwendet werden.

Die Erfindung betrifft außerdem ein Verfahren zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon.

Wird die Umsetzung mit einem rekombinanten Mikroorganismus durchgeführt, so erfolgt vorzugsweise zunächst die Kultivierung der Mikroorganismen in Gegenwart von Sauerstoff und in einem Komplexmedium, wie z.B. bei einer Kultivierungstemperatur von etwa 20 °C oder mehr, und einem pH-Wert von etwa 6 bis 9, bis eine ausreichende Zelldichte erreicht ist. Um die Reaktion besser steuern zu können, bevorzugt man die Verwendung eines induzierbaren Promotors. Die Kultivierung wird nach Induktion der Vitamin B2-Produktion in Gegenwart von Sauerstoff 12 Stunden bis 3 Tage fortgesetzt.

Folgende nichtlimitierende Beispiele beschreiben spezielle Ausführungsformen der Erfindung.

Allgemeine experimentelle Angaben

a) Allgemeine Klonierungsverfahren

Die im Rahmen der vorliegenden Erfindung durchgeführten Klonierungsschritte wie z.B. Restriktionsspaltungen, Agarose Gelelektrophorese, Reinigung von DNA-Fragmenten, Transfer von

Nukleinsäuren auf Nitrozellulose und Nylonmembranen, Verknüpfen von DNA-Fragmenten, Transformation von *E. coli* Zellen, Anzucht von Bakterien, Vermehrung von Phagen und Sequenzanalyse rekombinanter DNA wurden wie bei Sambrook et al. (1989) a.a.O. beschrieben durchgeführt.

5

b) Polymerasekettenreaktion (PCR)

PCR wurde nach Standardprotokoll mit folgendem Standardansatz durchgeführt:

- 10 8 µl dNTP-Mix (200µM), 10 µl Taq-Polymerase-Puffer (10 x) ohne MgCl₂, 8µl MgCl₂ (25mM), je 1 µl Primer (0,1 µM), 1µl zu amplifizierende DNA, 2,5 U Taq-Polymerase (MBI Fermentas, Vilnius, Litauen), ad 100 µl demineralisiertes Wasser.

c) Kultivierung von *E.coli*

15

Die Kultivierung von rekombinanten *E. coli*-Stämme DH5α wurde in LB-Amp Medium (Trypton 10,0g, NaCl 5,0 g, Hefeextrakt 5,0 g, Ampicillin 100 g/ml H₂O ad 1000 ml) bei 37 °C kultiviert. Dazu wurde jeweils eine Kolonie mittels Impföse von einer Agarplatte in 5 ml LB-Amp überführt. Nach ca. 18 h Stunden Kultivierung bei einer Schüttelfrequenz von 220 Upm wurden 400 ml Me-

20 dium in einem 2-l-Kolben mit 4 ml Kultur inokuliert. Die Induktion der P450-Expression in *E. coli* erfolgte nach Erreichen eines OD₅₇₈-Wertes zwischen 0,8 und 1,0 durch eine drei- bis vierstündige Hitzeschockinduktion bei 42 °C.

d) Reinigung des gewünschten Produktes aus der Kultur

25

Die Gewinnung des gewünschten Produktes aus dem Mikroorganismus oder aus dem Kulturüberstand kann durch verschiedene, im Fachgebiet bekannte Verfahren erfolgen. Wird das gewünschte Produkt von den Zellen nicht sezerniert, können die Zellen aus der Kultur durch langsame Zentrifugation geerntet werden, die Zellen können durch Standard-Techniken, wie mechanische Kraft oder Ultraschallbehandlung, lysiert werden.

30

Die Zelltrümmer werden durch Zentrifugation entfernt, und die Überstandsfraktion, die die löslichen Proteine enthält, wird zur weiteren Reinigung der gewünschten Verbindung erhalten. Wird das Produkt von den Zellen sezerniert, werden die Zellen durch langsame Zentrifugation aus der

35 Kultur entfernt, und die Überstandsfraktion wird zur weiteren Reinigung behalten.

Die Überstandsfraction aus beiden Reinigungsverfahren wird einer Chromatographie mit einem geeigneten Harz unterworfen, wobei das gewünschte Molekül mit höherer Selektivität als die Verunreinigungen entweder auf dem Chromatographieharz zurückgehalten wird oder dieses passiert. Diese Chromatographieschritte können nötigenfalls wiederholt werden, wobei die gleichen oder andere Chromatographieharze verwendet werden. Der Fachmann ist in der Auswahl der geeigneten Chromatographieharze und ihrer wirksamsten Anwendung für ein bestimmtes zu reinigendes Molekül bewandert. Das gereinigte Produkt kann durch Filtration oder Ultrafiltration konzentriert und bei einer Temperatur aufbewahrt werden, bei der die Stabilität des Produktes maximal ist.

Im Stand der Technik sind viele Reinigungsverfahren bekannt. Diese Reinigungstechniken sind z.B. beschrieben in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

Die Identität und Reinheit der isolierten Verbindungen kann durch Techniken des Standes der Technik bestimmt werden. Diese umfassen Hochleistungs-Flüssigkeitschromatographie (HPLC), spektroskopische Verfahren, Färbeverfahren, Dünnschichtchromatographie, NIRS, Enzymtest oder mikrobiologische Tests. Diese Analyseverfahren sind zusammengefaßt in: Patek et al. (1994) Appl. Environ. Microbiol. 60:133-140; Malakhova et al. (1996) Biotekhnologiya 11 27-32; und Schmidt et al. (1998) Bioprocess Engineer. 19:67-70. Ullmann's Encyclopedia of Industrial Chemistry (1996) Bd. A27, VCH: Weinheim, S. 89-90, S. 521-540, S. 540-547, S. 559-566, 575-581 und S. 581-587; Michal, G (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, Bd. 17.

e) Allgemeine Beschreibung der MPSS-Methode, Klonidentifizierung und Homologiesuche

Die MPSS Technologie (**M**assive **P**arallele **S**ignatur **S**equenzierung, wie von Brenner et al, Nat. Biotechnol.(2000) 18, 630-634 beschrieben; worauf hiermit ausdrücklich Bezug genommen wird) wurde an dem filamentösen, Vitamin B2 produzierenden Pilz *Ashbya gossypii* angewendet. Mit Hilfe dieser Technologie ist es möglich, mit hoher Genauigkeit quantitative Aussagen über die Expressionsstärke einer Vielzahl von Genen in einem eukaryotischen Organismus zu erhalten. Dabei wird die mRNA des Organismus zu einem bestimmten Zeitpunkt X isoliert, mit Hilfe des Enzyms Reverse Transkriptase in cDNA umgeschrieben und anschließend in spezielle Vektoren kloniert, die eine spezifische Tag-Sequenz besitzen. Die Anzahl von Vektoren mit unterschiedlicher Tagsequenz wird dabei so hoch gewählt (etwa 1000-fach höher), dass statistisch gesehen, jedes DNA-Molekül in einen, durch seine Tag-Sequenz einzigartigen, Vektor kloniert wird.

Anschließend werden die Vektorinserts zusammen mit dem Tag herausgeschnitten. Die so erhaltenen DNA-Moleküle werden dann mit Mikrokügelchen inkubiert, die die molekularen Gegenstücke zu den erwähnten Tags besitzen. Nach Inkubation kann davon ausgegangen werden, dass jedes Mikrokügelchen über die spezifischen Tags bzw. Gegenstücke mit nur einer Sorte von DNA Molekülen beladen ist. Die Kügelchen werden in eine spezielle Flußzelle überführt und dort fixiert, so dass es möglich ist, mit Hilfe eines adaptierten Sequenzierverfahrens, auf Basis von Fluoreszenzfarbstoffen und mit Hilfe einer digitalen Farbkamera, eine Massensequenzierung aller Kügelchen vorzunehmen. Mit dieser Methode ist zwar eine zahlenmäßig hohe Auswertung möglich, die allerdings durch eine Leseweite von etwa 16 bis 20 Basenpaaren limitiert ist. Die Sequenzlänge reicht dennoch aus, um bei den meisten Organismen eine eindeutige Zuordnung zwischen Sequenz und Gen zu ermöglichen (20bp besitzen eine Sequenzhäufigkeit von $\sim 1 \times 10^{12}$, das menschliche Genom besitzt im Vergleich dazu "nur" eine Größe von $\sim 3 \times 10^9$ bp).

Die auf diese Weise erhaltenen Daten werden ausgewertet, indem die Anzahl der gleichen Sequenzen gezählt und ihre Häufigkeiten miteinander verglichen werden. Häufig auftretende Sequenzen spiegeln eine hohe Expressionsstärke, vereinzelt auftretende Sequenzen eine niedrige Expressionsstärke wider. Erfolgte die mRNA-Isolation zu zwei Unterschiedlichen Zeitpunkten (X und Y), so ist es möglich ein zeitliches Expressionsmuster einzelner Gene aufzustellen.

Beispiel 1:

Isolation von mRNA aus *Ashbya gossypii*

Ashbya gossypii wurde in an sich bekannter Weise kultiviert (Nährmedium: 27,5 g/l Hefeextrakt; 0,5 g/l Magnesiumsulfat; 50ml/l Sojaöl; pH 7). Myzelproben von *Ashbya gossypii* werden zu unterschiedlichen Zeitpunkten der Fermentation (24h, 48h und 72h) entnommen und die entsprechende RNA bzw. mRNA wird nach dem Protokoll von Sambrook et al. (1989) daraus isoliert.

Beispiel 2:

Anwendung der MPSS

Isolierte mRNA von *A. gossypii* wird dann einer MPSS-Analyse, wie oben erläutert, unterzogen.

Die ermittelten Datensätze werden einer statistischen Auswertung unterzogen und nach Signifikanz der Expressionsunterschiede gegliedert. Dabei wurde sowohl hinsichtlich Erhöhung bzw. Erniedrigung der Expressionsstärke untersucht. Eine Einteilung erfolgt über eine Einstufung der

Expressionsveränderung in a) monotone Veränderung, b) Veränderung nach 24h, und c) Veränderung nach 48h.

- 5 Die eine Expressionsveränderung repräsentierenden durch MPSS-Analyse ermittelten 20bp-Sequenzen werden dann als Sonden verwendet und gegen eine Genbank von *Ashbya gossypii*, mit einer durchschnittlichen Insertgröße von etwa 1kb, hybridisiert. Die Hybridisierungstemperatur lag dabei im Bereich von etwa 30 bis 57°C.

Beispiel 3:

10 Erstellung einer genomischen Genbank aus *Ashbya gossypii*

Zur Erstellung einer genomischen DNA-Bank wird zunächst chromosomale DNA nach der Methode von Wright und Philippsen (Gene (1991) 109: 99-105) und Mohr (1995, PhD Thesis, Biozentrum Universität Basel, Schweiz) isoliert.

15

Die DNA wird partiell mit Sau3A verdaut. Dazu werden 6µg genomische DNA einem Sau3A Verdau mit unterschiedlichen Enzymmengen (0,1 bis 1 U) unterzogen. Die Fragmente werden in einem Saccharose-Dichtegradienten fraktioniert. Die 1kb Region wird isoliert und einer QiaEx-Extraktion unterzogen. Die größten Fragmente werden mit dem BamHI geschnittenen Vektor pRS416 (Sikorski und Hieter, Genetics (1988) 122: 19-27) ligiert (90 ng BamHI geschnittener, dephosphorylierter Vektor; 198 ng Insert DNA; 5ml Wasser; 2 µl 10xLigationspuffer; 1U Ligase). Mit diesem Ligationsansatz wird *E. coli* Laborstamm XL-1 blue transformiert und die resultierenden Klone werden zur Identifizierung des Inserts eingesetzt.

20

25 Beispiel 4:

Herstellung einer geordneten Genbank (CHIP-Technologie)

30

Etwa 25,000 Kolonien der *Ashbya gossypii* Genbank (dies entspricht einer etwa 3-fachen Genomabdeckung) wurden geordnet auf eine Nylonmembran transferiert und anschließend nach der Methode der Koloniehybridisierung wie in Sambrook et al. (1989) beschrieben, behandelt. Von den durch MPSS-Analyse ermittelten 20bp-Sequenzen wurden Oligonukleotide synthetisiert und mit Hilfe von ³²P radioaktiv markiert. Jeweils 10 markierte Oligonukleotide mit ähnlichem Schmelzpunkt werden vereinigt und gemeinsam gegen die Nylonmembranen hybridisiert. Nach Hybridisierungs- und Waschschritten werden positive Klone durch Autoradiographie identifiziert, und mit Hilfe von PCR-Sequenzierung direkt analysiert.

35

Auf diese Weise wurde ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 19“ trägt und mit dem MIPS Tag „Ygr257c“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 1.

- 5 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 24“ trägt und mit dem MIPS Tag „Mdl2“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 5.

- 10 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 109“ trägt und mit dem MIPS Tag „Prp12“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 10.

- 15 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 163“ trägt und mit dem MIPS Tag „Flx1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 14.

- 20 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 31“ trägt und mit dem MIPS Tag „STV1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 19.

Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 4“ trägt und mit dem MIPS Tag „OPT2“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 23.

- 25 Auf diese Weise wurde ein weiterer Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 6“ trägt und mit dem MIPS Tag „VAC1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 28.

- 30 Auf diese Weise wurde ein weiterer Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 146“ trägt und mit dem MIPS Tag „Ymr162c“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 33.

- 35 Auf diese Weise wurde ein weiterer Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 56“ trägt und mit dem MIPS Tag „Ypl110c“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 37.

Auf diese Weise wurde einweiterer Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 167“ trägt und mit dem MIPS Tag „ERP5“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 42.

5 Beispiel 5:

Auswertung der Sequenzdaten mit Hilfe einer BLASTX Suche

Eine Auswertung der erhaltenen Nukleinsäure-Sequenzen, d.h. deren funktionale Zuordnung zu einer funktionalen Aminosäuresequenz, erfolgte mittels einer BLASTX-Suche in Sequenz-

10 Datenbanken. Fast alle der aufgefundenen Aminosäuresequenz-Homologien betrafen *Saccharomyces cerevisiae* (Bäckerhefe). Da dieser Organismus bereits vollständig sequenziert worden ist, konnten genauere Informationen bezüglich dieser Gene unter:

http://www.mips.gsf.de/proj/yeast/search/code_search.htm nachgeschlagen werden.

15 So wurden folgende Homologien mit einem Aminosäurefragment aus *S. cerevisiae* ermittelt. Das entsprechenden Alignment sind in den beiliegenden Figur 1 bis 10 gezeigt.

a) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:1 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem mitochondrialen energie-

20 übertragenden Protein aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 609 bis 1 aus SEQ ID NO:1) mit einer Teilsequenz des *S. cerevisiae* Proteins ist in Figur 1 dargestellt. SEQ ID NO: 2 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

25 Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines mitochondrialen energieübertragenden Proteins zugeordnet werden.

b) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:5 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem ABC-Transportprotein aus *S.*

30 *cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (SEQ ID NO:6) (entsprechend den Nucleotiden 1494 bis 1387 aus SEQ ID NO:5) mit einer Teilsequenz des *S. cerevisiae* Proteins ist in Figur 2 dargestellt. SEQ ID NO: 7 zeigt eine weitere erfindungsgemäße Aminosäure-Teilsequenz.

35 Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines ABC-Transportproteins zugeordnet werden.

- c) Die vom kodierenden Strang zu SEQ ID NO:10 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem membranintegrierten mitochondrialen Protein aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 15 bis 455 aus SEQ ID NO:10) mit einer Teilsequenz des *S. cerevisiae* Proteins ist in Figur 3 dargestellt. SEQ ID NO: 11 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines membranintegrierten mitochondrialen Proteins zugeordnet werden.

- d) Die vom kodierenden Strang zu SEQ ID NO:14 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem mitochondrialen Innermembran-Transportprotein aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 415 bis 1215 aus SEQ ID NO:14) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 4 dargestellt. SEQ ID NO: 15 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines mitochondrialen Innermembran-Transportproteins zugeordnet werden.

- e) Die vom kodierenden Strang zu SEQ ID NO:19 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer nicht-vakuolären 102 kD-Untereinheit der H⁺-ATPase-VO-Domäne aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 2 bis 790 aus SEQ ID NO:19) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 5 dargestellt. SEQ ID NO: 20 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer nicht-vakuolären 102 kD-Untereinheit der H⁺-ATPase-VO-Domäne zugeordnet werden.

- f) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:23 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem Protein aus *S. cerevisiae* mit einer Ähnlichkeit zum isp4-Protein aus *S. pombe*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 869 bis 522 aus SEQ ID NO:23) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 6 dargestellt. SEQ ID NO: 24 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines Proteins mit einer Ähnlichkeit zum isp4-Protein aus *S. pombe* und damit der Aktivität eines Oligopeptid-Transporters zugeordnet werden.

- 5 g) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:28 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem VAC1-Protein, einem cytosolischen und peripheren Membranprotein mit drei Zinkfingern, aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 356 bis 243 aus SEQ ID NO:28) mit
10 abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 166 bis 2 aus SEQ ID NO:28) mit einer Teilsequenz des *S. cerevisiae* Proteins ist in Figur 7A dargestellt. Eine weitere davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 166 bis 2 aus SEQ ID NO:28) mit einer Teilsequenz des *S. cerevisiae* Proteins ist in Figur 7B gezeigt. SEQ ID NO: 29 und SEQ ID NO: 30 zeigen jeweils eine N-terminal verlängerte Aminosäure-Teilsequenz .

15 Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines VAC1-Proteins, eines cytosolischen und peripheren Membranproteins mit drei Zinkfingern, zugeordnet werden.

- h) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:33 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem Protein mit einer ATPase-ähnlichen Funktion aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend
20 den Nucleotiden 904 bis 707 aus SEQ ID NO:33) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 8 dargestellt. SEQ ID NO: 34 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

25 Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines ATPase-ähnlichen Proteins zugeordnet werden.

- i) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:37 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem PHO85-Protein aus *S. cerevisiae*. Eine davon abgeleitete Aminosäure-Teilsequenz (entsprechend den Nucleotiden 898 bis 5 aus
30 SEQ ID NO:37) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 9 dargestellt. Die Aminosäuresequenzen gemäß SEQ ID NO: 38 und SEQ ID NO: 39 entsprechen Aminosäure-Teilsequenzen abgeleitet vom Gegenstrang zu Position 950 bis 900 bzw. 898 bis 5 in SEQ ID NO:37.

35 Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines PHO85-Proteins zugeordnet werden.

- k) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:42 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit dem am Membranverkehr beteiligten p24-Protein aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 931 bis 806 aus SEQ ID NO:42) mit einer Teilsequenz des *S. cerevisiae* Proteins ist in Figur 10 dargestellt. SEQ ID NO: 42 zeigt jeweils eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines am Membranverkehr beteiligten p24-Proteins zugeordnet werden.

Beispiel 6:

Isolierung der Full-Length-DNA

- a) Konstruktion einer *A. gossypii*-Genbank

Hochmolekulare zelluläre Gesamt-DNA von *A. gossypii* wurde aus einer 2 Tage alten, in einem flüssigen MA2-Medium (10g Glucose, 10g Pepton, 1g Hefeextrakt, 0,3g Myo-Inosit ad 1000 ml) gewachsenen 100 ml Kultur hergestellt. Das Myzel wurde abfiltriert, zweimal mit H₂O dest. gewaschen, in 10 ml 1M Sorbitol, 20 mM EDTA, enthaltend 20 mg Zymolyase-20T, suspendiert und 30 bis 60 min unter leichtem Schütteln bei 27 °C inkubiert. Die Protoplasten-Suspension wurde auf 50 mM Tris-HCl, pH 7,5, 150 mM NaCl, 100 mM EDTA und 0,5-%igem Natriumdodecylsulfat (SDS) eingestellt und 20 min bei 65 °C inkubiert. Nach zwei Extraktionen mit Phenol-Chloroform (1:1 vol/vol) wurde die DNA mit Isopropanol gefällt, in TE-Puffer suspendiert, mit RNase behandelt, erneut mit Isopropanol gefällt und in TE resuspendiert.

Eine *A. gossypii*-Kosmid-Genbank wurde hergestellt, indem man nach der Größe ausgewählte, mit *Sau3A* teilverdaute genomische DNA an die dephosphorylierten Arme des Cosmidvektors Super-Cos1 (Stratagene) band. Der Super-Cos1-Vektor wurde zwischen den beiden *cos*-Stellen durch Verdau mit *Xba*I und Dephosphorylierung mit alkalischer Kalbsdarm-Phosphatase (Boehringer) geöffnet, gefolgt von einem Öffnen der Klonierungsstelle mit *Bam*HI. Die Ligationen wurden über Nacht bei 15 °C in 20 µl, enthaltend 2,5 µg teilverdauter chromosomaler DNA, 1 µg Super-Cos1-Vektorenarme, 40 mM Tris-HCl, pH 7,5, 10 mM MgCl₂, 1mM Dithiothreitol, 0,5 mM ATP und 2 Weiss-Einheiten T4-DNA-Ligase (Boehringer) durchgeführt. Die Ligationsprodukte wurden unter Verwendung der Extrakte und des Protokolls von Stratagene (Gigapack II Packaging Extract) in vitro verpackt. Das verpackte Material wurde zur Infizierung von *E. coli* NM554 (*recA13*, *araD139*, Δ (*ara*,*leu*)7696, Δ (*lac*)17A, *galU*, *galK*, *hsrR*, *rps*(*str*'), *mcrA*, *mcrB*) verwendet und auf Ampicillin (50 µg/ml) enthaltende LB-Platten verteilt. Man erhielt Transformanten, welche

ein *A. gossypii*-Insert einer durchschnittlichen Länge von 30-45 kb enthielten.

b) Lagerung und Screening der Cosmid-Genbank

- 5 Insgesamt 4×10^4 frische Einzelkolonien wurden einzeln in Vertiefungen von 96-er Microtiterplatten (Falcon, Nr. 3072) in 100 µl LB-Medium, ergänzt mit dem Gefriermedium (36 mM K_2HPO_4 /13,2 mM KH_2PO_4 , 1,7 mM Natriumcitrat, 0,4 mM $MgSO_4$, 6,8 mM $(NH_4)_2SO_4$, 4,4% (wt/vol) Glycerin) und Ampicillin (50 µg/ml), inokuliert, über Nacht bei 37 °C unter Schütteln wachsen gelassen und bei -70 °C eingefroren. Die Platten wurden rasch aufgetaut und danach
- 10 unter Verwendung eines 96-er-Replikators, der in einem Ethanolbad unter anschließender Verdunstung des Ethanols auf einer heißen Platte sterilisiert worden war, in frisches Medium dupliziert. Vor dem Einfrieren und nach dem Auftauen (vor irgendwelchen anderen Maßnahmen) wurden die Platten kurz in einem Mikrotiterschüttler (Infors) geschüttelt, um eine homogene Zellsuspension zu gewährleisten. Mittels eines Robotersystems (Bio-Robotics), mit dem geringe
- 15 Mengen an Flüssigkeit aus 96 Vertiefungen einer Mikrotiterplatte auf Nylonmembran (GeneScreen Plus, New England Nuclear) transferiert werden können, wurden einzelne Klone auf Nylonmembranen platziert. Nach dem Transfer der Kultur aus den 96-er Mikrotiterplatten (1920 Klone) wurden die Membranen auf die Oberfläche von LB-Agar mit Ampicillin (50 µg/ml) in 22 x 22 cm Kulturschalen (Nunc) platziert und über Nacht bei 37 °C inkubiert. Vor Erreichen der Zell-
- 20 konfluenz wurden die Membranen, wie von Herrmann, B. G., Barlow, D. P. und Lehrach, H. (1987) in Cell 48, S. 813-825 beschrieben, prozessiert, wobei als zusätzliche Behandlung nach dem ersten Denaturierungsschritt ein 5-minütiges Bedampfen der Filter auf einem in Denaturierungslösung getränkten Pad über einem kochenden Wasserbad hinzukommt.
- 25 Hit Hilfe des Random-Hexamer-Primer-Verfahrens (Feinberg, A. P. und Vogelstein, B. (1983), Anal. Biochem. 132, S. 6-13) wurden doppelsträngige Sonden durch Aufnahme von [α - ^{32}P]dCTP mit hoher spezifischer Aktivität markiert. Die Membranen wurden prähybridisiert und 6 bis 12 h bei 42 °C in 50% (vol/vol) Formamid, 600 mM Natriumphosphat, pH 7,2, 1 mM EDTA, 10% Dextransulfat, 1% SDS, und 10x Denhardt-Lösung, enthaltend Lachssperma-DNA (50 µg/ml) mit ^{32}P -markierten Sonden ($0,5-1 \times 10^6$ cpm/ml) hybridisiert. Typischerweise wurden
- 30 Waschschrte etwa 1 h bei 55 bis 65 °C in 13 bis 30 mM NaCl, 1,5 bis 3 mM Natriumcitrat, pH 6,3, 0,1 % SDS durchgeführt und die Filter wurden 12 bis 24 h bei -70 °C mit Kodak-Verstärkerplatten autoradiographiert. Bislang wurden einzelne Membrane mehr als 20 mal erfolgreich wiederverwendet. Zwischen den Autoradiographien wurden die Filter durch Inkubation
- 35 bei 95 °C für 2 x 20 min in 2 mM Tris-HCl, pH 8,0, 0,2 mM EDTA, 0,1 % SDS gestrippt.

c) Zurückgewinnung positiver Kolonien aus der aufbewahrten Genbank

Gefrorene Bakterienkulturen in Microtiter-Wells wurden unter Verwendung steriler Einweg-Lanzetten abgekratzt und das Material wurde auf LB-Agar-Petrischalen enthaltend Ampicillin (50 µg/ml) ausgestrichen. Einzelne Kolonien wurden danach zur Inokulierung von Flüssigkulturen für
5 Herstellung von DNA mittels Alkali-Lyse-Verfahren (Birnboim, H. C. und Doly, J. (1979), Nucleic Acids Res. 7, S. 1513-1523) verwendet.

d) Full-Length DNA

10 Auf die oben beschriebenen Weise konnten Klone identifiziert werden, der ein Insert mit der entsprechenden Vollsequenz tragen. Diese Klone tragen die internen Bezeichnungen:

„Oligo 19v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß
SEQ ID NO: 3.

15

„Oligo 24v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß
SEQ ID NO: 8.

Oligo 109v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß
20 SEQ ID NO: 12.

„Oligo 163v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß
SEQ ID NO: 17.

25 „Oligo 31v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß
SEQ ID NO: 21.

„Oligo 4v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß
SEQ ID NO: 26.

30

„Oligo 6v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß
SEQ ID NO: 31.

35 „Oligo 146v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz ge-
mäß SEQ ID NO: 35.

„Oligo 56v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz ge-

gemäß SEQ ID NO: 40.

„Oligo 167v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 44.

5

Eine Übersicht über sämtliche erfindungsgemäße Teil- und Vollsequenzen ist in folgender Tabelle 1 zu finden:

Tabelle 1: Sequenzübersicht

5

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
1	019	DNA-Teilsequenz	mitochondriales energieübertragendes Protein aus <i>S. cerevisiae</i>
2	019	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 1	
3	019	DNA-Full-lenght-Sequenz	
4	019	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 3 von Position 112 bis 294	
5	024	DNA-Teilsequenz	ABC-Transportproteins aus <i>S. cerevisiae</i>
6	024	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 5	
7	024	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 5	
8	024	DNA-Full-lenght-Sequenz	
9	024	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 8 von Position 820 bis 3081	
10	109	DNA-Teilsequenz	membranintegriertes mitochondriales Protein aus <i>S. cerevisiae</i> .
11	109	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 10	
12	109	DNA-Full-lenght-Sequenz	
13	109	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 12 von Position 502 bis 2919	
14	163	DNA-Teilsequenz	mitochondriales Innermembran-Transportprotein aus <i>S. cerevisiae</i>
15	163	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 14	
16	163	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 14	
17	163	DNA-Full-lenght-Sequenz	

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
18	163	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 17 von Position 329 bis 1207	
19	31	DNA-Teilsequenz	nicht-vakuoläre 102 kD-Untereinheit der H ⁺ -ATPase-VO-Domäne aus <i>S. cerevisiae</i> .
20	31	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 19	
21	31	DNA-Full-lenght-Sequenz	
22	31	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 21 von Position 623 bis 3253	
23	4	DNA-Teilsequenz	isp4-Protein aus <i>S. pombe</i> Protein mit vergleichbarer Funktion aus <i>S. cerevisiae</i> .
24	4	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 23	
25	4	DNA-Full-lenght-Sequenz (Gegenstrang)	
26	4	DNA-Full-lenght-Sequenz mit ORF-Bereich	
27	4	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 26 von Position 738 bis 1037	
28	6	DNA-Teilsequenz	VAC1-Protein, ein cytosolisches und peripheres Membranprotein mit drei Zinkfingern aus <i>S. cerevisiae</i>
29	6	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 28	
30	6	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 28	
31	6	DNA-Full-lenght-Sequenz	
32	6	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 31 von Position 428 bis 1993	
33	146	DNA-Teilsequenz	Protein mit einer ATPase oder ATPase-ähnlichen Funktion aus <i>S. cerevisiae</i>
34	146	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 33	
35	146	DNA-Full-lenght-Sequenz	

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
36	146	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 35 von Position 537 bis 1034	
37	56	DNA-Teilsequenz	PHO85-Protein aus <i>S. cerevisiae</i>
38	56	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 37	
39	56	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 37	
40	56	DNA-Full-length-Sequenz	
41	56	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 40 von Position 426 bis 4388	
42	167	DNA-Teilsequenz	p24-Protein aus <i>S. cerevisiae</i>
43	167	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 42	
44	167	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 42	
45	167	DNA-Full-length-Sequenz	
46	167	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 45 von Position 563 bis 1216	

Patentansprüche

1. Polynukleotid, welches aus *Ashbya gossypii* isolierbar ist und für ein Protein kodiert, das mit dem Transmembrantransport assoziiert ist und/oder ein Transmembranprotein ist.
- 5 2. Polynukleotid, nach Anspruch 1, das mit dem Transmembrantransport assoziiert ist und/oder ein Transmembranprotein ist und eine in Tabelle 1 angegebene strukturelle und/oder funktionale Eigenschaft besitzt.
- 10 3. Polynukleotid nach Anspruch 1 oder 2, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 1, 5, 10, 14, 19, 23, 28, 33, 37 oder 42, welche vorzugsweise aus *Ashbya gossypii* isolierbar ist; das dazu komplementäre Polynukleotid; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.
- 15 4. Polynukleotid nach Anspruch 4, welches eine Nukleinsäuresequenz gemäß SEQ ID NO: 3, 8, 12, 17, 21, 26, 31, 35, 40 oder 44 oder ein Fragment davon umfasst.
5. Oligonukleotid, welches mit einem Polynukleotid nach einem der vorhergehenden Ansprüche, insbesondere unter stringenten Bedingungen, hybridisiert.
- 20 6. Polynukleotid, welches mit einem Oligonukleotid nach Anspruch 5, insbesondere unter stringenten Bedingungen, hybridisiert und für ein Genprodukt aus Mikroorganismen der Gattung *Ashbya* oder ein funktionales Äquivalent dieses Genproduktes kodiert.
- 25 7. Polypeptid, welches von einem Polynukleotid kodiert wird, das die Nukleinsäuresequenz gemäß Anspruch 1 bis 4 oder ein Fragment davon umfasst, oder von einem Polynukleotid nach Anspruch 6 kodiert wird; oder welches eine Aminosäuresequenz aufweist, die wenigstens 10 zusammenhängende Aminosäurereste gemäß SEQ ID NO: 2, 4, 6, 7, 9, 11, 13, 15, 16, 18, 20, 22, 24, 27, 29, 30, 32, 34, 36, 38, 39, 41, 43 oder SEQ ID NO: 45 umfasst; sowie funktionale Äquivalente davon, insbesondere solche welche eine der in
- 30 8. Expressionskassette, umfassend in operativer Verknüpfung mit wenigstens einer regulativen Nukleinsäuresequenz eine Nukleinsäuresequenz nach einem der Ansprüche 1 bis 6.
- 35 9. Rekombinanter Vektor, umfassend wenigstens eine Expressionskassette nach Anspruch 8.

10. Prokaryotischer oder eukaryotischer Wirt, transformiert mit wenigstens einem Vektor nach Anspruch 9.
- 5 11. Prokaryotischer oder eukaryotischer Wirt, in welchem die funktionale Expression wenigstens eines Gens moduliert ist, das für ein Polypeptid nach Anspruch 7 kodiert; oder in welchem die biologische Aktivität eines Polypeptids nach Anspruch 7 erniedrigt oder erhöht ist.
- 10 12. Wirt nach Anspruch 10 oder 11 aus der Gattung *Ashbya*.
13. Verwendung einer Expressionskassette nach Anspruch 8, eines Vektors nach Anspruch 9 oder eines Wirts nach einem der Ansprüche 10 bis 12 zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon.
- 15 14. Verwendung einer Expressionskassette nach Anspruch 8, eines Vektors nach Anspruch 7 oder eines Wirts nach einem der Ansprüche 10 bis 12 zur rekombinanten Herstellung eines Polypeptids nach Anspruch 7.
- 20 15. Verfahren zum Nachweis eines Effektortargets für die Modulation der mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon, wobei man einen Mikroorganismus, der zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon befähigt ist, mit einem Effektor behandelt, welcher mit einem Target, ausgewählt unter einem Polypeptid nach Anspruch 7 oder einer
- 25 dafür kodierenden Nukleinsäuresequenz, wechselwirkt, insbesondere bindet, den Einfluß des Effektors auf die Menge des mikrobiologisch produzierten Vitamins B2, und/oder des Präkursors und/oder eines Derivats davon validiert; und das Target gegebenenfalls isoliert.
- 30 16. Verfahren zur Modulation der mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon, wobei man einen Mikroorganismus, der zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon befähigt ist, mit einem Effektor behandelt, welcher mit einem Target, ausgewählt unter einem Polypeptid nach Anspruch 7 oder einer dafür kodierenden Nukleinsäuresequenz, wechselwirkt.
- 35 17. Effektor für ein Target, ausgewählt unter einem Polypeptid nach Anspruch 7 oder einer dafür kodierenden Nukleinsäuresequenz, wobei der Effektor ausgewählt ist unter:

- 5
- a) Antikörpern oder antigenbindenden Fragmenten davon;
 - b) von a) verschiedenen Polypeptid-Liganden, welche mit dem Polypeptid gemäß Anspruch 7 wechselwirken;
 - c) niedermolekularen Effektoren, welche die biologische Aktivität eines Polypeptids nach Anspruch 7 modulieren;
 - d) Antisense-Nukleinsäuresequenzen.
- 10
18. Verfahren zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon, wobei man einen Wirt nach einem der Ansprüche 10 bis 12 unter die Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon begünstigenden Bedingungen kultiviert und das(die) gewünschte(n) Produkt(e) aus dem Kulturansatz isoliert.
- 15
19. Verfahren nach Anspruch 18, wobei man vor und/oder während der Kultivierung des Wirts diesen mit einem Effektor nach Anspruch 17 behandelt.
- 20
20. Verfahren nach Anspruch 18 oder 19, wobei der Wirt ausgewählt ist unter Mikroorganismen der Gattung *Ashbya*.
- 20
21. Verfahren nach einem der Ansprüche 18 bis 20, wobei der Mikroorganismus ein Wirt nach einem der Ansprüche 10 bis 12 ist.
- 25
22. Verwendung eines Polynukleotids nach einem der Ansprüche 1 bis 4 und 6 oder eines Polypeptids nach Anspruch 7 als Target zur Modulation der Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon in einem Mikroorganismus der Gattung *Ashbya*.
- 30
23. Verwendung eines Polynukleotids nach einem der Ansprüche 1 bis 4 und 6 oder eines Polypeptids nach Anspruch 7 als Target zur Modulation des Transmembran-Transports, der Aktivität eines Transmembranproteins oder eines damit zusammenhängenden Folgezustandes in einem Mikroorganismus der Gattung *Ashbya* während der Kultivierung zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon.
- 35
24. Wirt nach Anspruch 12 mit verbesserter Zellantwort auf äußere Bedingungen.

Fig. 1

Oligo 19

609 KEPGKKVTLKXRMVSASAGSLVTSFLTPLDVVRVRLQQQEMLPSCTCTGQLSKPAG--- 439

+ +TLK RM+SA AGS++TSL LTP+DVVR+RLQQQ+M+P C+C G P

4 RNTSNSLTALKERMLSAGAGSVLTSILTPMDVVRIRLQQQQMIPDCSCDGAAEVPNAVSS 63

438 -----KVFWDQECFANVGCREPAARLQGTLEGLRKIAQLEGLPTLWR 313

K+FW+ CF + C+ + + GTLE KIA +EG+ +LWR

64 GSKMKTFTNVGGQNLNNAKIFWESACFQELHCKNSSLKFNGTLEAFTKIASVEGITSWLWR 123

312 GLGITLVMAVPANVVYFSGYEALRDNSPLASRLPVANPLVCGAFARILAATTIAPLELLR 133

G+ +TL+MA+PAN+VYFSGYE +RD SP+AS P NPL CGA AR+ AAT+IAPLEL++

124 GISLTLLMAIPANMVYFSGYEYIRDVSPIASTYPTLNPLFCGAIARVFAATSIAPLELVK 183

132 TRLQSVPRARDTERTIYLLIGDLLREMRHEVSVMG-YRALFKGLEI 1 A. gossypii

T+LQS+PR+ + +T ++ DLL E R E+ ++G RALFKGLEI

184 TKLQSI PRSSKSTKTWMMVKDLLNETRQEMKMVGPSRALFKGLEI 228 S. cerevisiae

Fig. 2**Oligo 24**

1494 ILKKLRLTITPGSNVCIVGPSGRGKSTIASLLLRFY 1387 A. gossypii
I K L I PGS+VCIVGPSGRGKSTIA LLLR+Y
510 IFKNLNFKIAPGSSVCIVGPSGRGKSTIALLLLRYY 545 S. cerevisiae

Fig. 3**Oligo109**

```
15   KTAQAWELIELLSQNDVVKYGDIVFRPLFKSSPEAGLLELEKNGLITISRNRGVLQDIRP 194
      K+AQAWELIELLS N V+ + +IV +PLFK++PE G++ELE NGLIT+SR+RGVLQ+IRP
704   KSAQAWELIELLSANPVIPIFHEIVNKPLFKAAPETGIMELENGLITVSRDRGVLQEIRP 763

195   AKPLFKAAFSYLLQDKDLSIVLRTGYLLRLIAFETGRIKKWEEELRLLAKVSDQRICKSR 374
      AKPL++AAF+YL+ D +L+ VL+T Y L+++ FETGRIKKWEEEL+ L KV DQ++ K+R
764   AKPLYRAAFTYLINDPELAKVLKTRYLLKVVGFTGRIKKWEEELKPLGKVPDQKLFKTR 823

375   LNYLASKIDASSGVINSCEDKVKEMSK 455   A. gossypii
      L+YL+ KI+AS+ VI CE+++K +SK
824   LDYLSGKINASNAVITKCEEEIKNLSK 850   S. cerevisiae
```

Fig. 4

Oligo163

415 HELTSLQREVISGLTAGTITTTIASHPLDLLKLRLQLSAGNRANTTYTG--LIRDIFERQ-
 H+ T LQ+EVISGL+AG++TT+ HPLDLLK+RLQLSA + Y +I++I
 4 HQWTP LQKEVISGLSAGSVTTLVVHPLDLLKVRLQLSATSQAQKAHYGPFMVIKEIIRSSA 63

 QWGR----ELYRGLGVNLLGNSVAWALYFGCYRCAKDIALRH LGNESATGIM----DRRL
 GR ELYRGL +NL GN++AW +YFG Y K++ + + T + D ++
 64 NSGRSVTNELYRGLSINLFGNAIAWGVYFGLYGVTKELIYKSVAKPGETQLKGVGNDHKM 123

 PAHAYMLAAGSSGIATAVLTNPIWVIKTRIMATSRA--GPYKSTFDGVYKLYQTEGVLA F
 + Y+ A SSG+ TA+LTNPIWVIKTRIM+TS+ G Y S ++GV +L +T+G
 124 NSLIYLSAGASSGLMTAILTNPIWVIKTRIMSTSKGAQGAYTSMYNGVQQLLR TDGFQGL 183

 WRGVVPSLLGVSQGAIYFALYDTLKFHYLHSSTDKA-ERRLSVSEIIGITCISKMISVTS
 W+G+VP+L GVSQGA+YFA+YDTLK L + + L+ E I IT + KM+SVT
 184 WKGLVPALFGVSQGALYFAVYDTLKQRKLRRKRENGLDIHLTNLETIEITS LGKMVS VTL 243

 VYPFQLLKSKLQDFGAPS---GITQLVQTVYSREGIRGFYRGLS 1215 A. gossypii
 VYPFQLLKS LQ F A + L++ + + +G G Y+GLS
 244 VYPFQLLKSNLQSFRA NEQKFRLFPLIKLIIANDGFVGLYKGLS S. cerevisiae

Fig. 5

Oligo 31

2 IDIVGNFIPGLIFMQSIFGYLSWAIYKWSKDWDKDELPAPGLLNMLINMFLSPGVVDEK 181
 +DI+GNFIPGL+FMQSIFGYLSWAI+YKWSKDWDK+ PAPGLLNMLINMFL+PG +D++
 612 VDIIGNFIPGLVFMQSIFGYLSWAIYKWSKDWDKDDKPAPGLLNMLINMFLAPGTIDDQ 671

 182 LYTGQSFLQVILLLAALVCVPWLLLYKPLMLKRQNDIA---LSKGFRSLRDQRVHEILLE 352
 LY+GQ+ LQV+LLLAALVCVPWLLLYKPL L+R N G++S+ + E + +
 672 LYSGQAKLQVVLLLAALVCVPWLLLYKPLTLRRLNKNNGGGRPHGYQSVGNIEHEEQIAQ 731

 353 AQENAG--EDMLVAXXXXXXXXXXXXXXG-----DVMIHQVIHTIEFCLNCISHTA 499
 + +A + M+++ G DVMIHQVIHTIEFCLNCISHTA
 732 QRHSAEGFQGMIIISDVASVADSINESVGGGEQGPFNFGDVMIHQVIHTIEFCLNCISHTA 791

 500 SYLRLWALSLAHAQLSTVLWSMTIQNSFSDSNPGSFFSVTKVVVLFAMWFVLTVCILVLM 679
 SYLRLWALSLAHAQLS+VLW MTI N+FS N GS +V KVV LFAMWFVLTVCILV M
 792 SYLRLWALSLAHAQLSSVLWDMTISNAFSSKNSGSPLAVMKVVFLFAMWFVLTVCILVFM 851

 680 EGTSAMLHSLRLHWVEAMSKFFELEGYAYEPFSFKAI 790 *A. gossypii*
 EGTSAMLH+LRLHWVEAMSKFFELEGYAYEPFSF+AI
 852 EGTSAMLHALRLHWVEAMSKFFELEGYAYEPFSFRAI 888 *S. cerevisiae*

Fig. 6**Oligo 4**

869 P*QRFKFTCPGIRKYVNTSFM*DSIDPKRISEHFYPIMKWCWLIVACEGLFFGLWKKLSL 690
P Q KFTCP Y N S + +I PKRI + YPI KWCWLI AC G+FFG+WK+
698 PHQNAKFTCPDAVTTYNASVVWGAIGPKRIFNYIYPIFKWCWLIGACIGIFFGVWKRWGK 757

689 MSEIYLILTLGLVELFNNGMIIAVSSYTFQYNTSGLCVGFINMLYMRRYYL-IWEKFN 522 (i)
Y LF GM+ Y Y TSG+ V +I+ YM+R++L +WEK+N
758 FYPRYFD----PMLFVGGMLNMSPPYNLMYYTSGMIVSYISQYYMKRHHLNLWEKYN 810 (ii)

(i) A. gossypii

(ii) S. cerevisiae

Fig. 7A**Oligo 6**

```
356 ESVECPVCNRNIENLEQLNMHLDVEHNFKD-ERSL-STDS    A. gossypii
      E+V CP+C R  +NL+ LN HLDVEH F D E SL    ++
4    ENVSCPICLRKFDNLQALNAHLDVEHGFNDNEDSLGSNDS    S. cerevisiae
```

Fig. 7B

```
166 RSHYKPLVPGKSQCSRCGRKLTTPRTGIQNCRRCGDLFCDIHCRLPVRLDLQGEYD 2    A. gossypii
      RSH++      GKS C  CGR L      G  NCR+CG L+C  H      ++L+L  +YD
65  RSHWEKFKKGKSCCHTCGRTLNNNIGAINCRKCGKLYCRRHLPNMIKLNLSAQYD 119 S. cerevisiae
```

Fig. 8**Oligo 146**

904 SSATYNTEEEYQLPSGKLIKRNLPDMSGGSGNDGFAAKLGRKLKL---KPAAEDIDEI 734
+S+ ++E YE LPSGKLIK R +L +S S KL +KLKL ED+++I
1588 NSSKKSSERYEVLPSGKLIKRP SLKTQSSKDSIGGNITTKLTKKLKLPSRNVEDEDVNQI 1647

733 IERRMQNLE 707 A. gossypii
I+ R+++LE
1648 IQARLKDLE 1656 S. cerevisiae

Fig. 9

Oligo 56

898 SSVVLIPTRTLVTLNKSALSRSFSCVDEILGIIPTLGDPDINGRNFFHHHVIALGK-- 725
 S + LIPTR ++++LNK+ALS+S +C+D IL +IP+L D DIN RNFFHHH+IA+GK
 292 SELSLIPTRIMISVLNKAALSKSLACIDAILKVIPSLNDSIEDINRRNFFHHHIIAIGKLI 351

 724 KRTKSLEERDNINXXXXXXXXXLEAAIPPEP-NTRLVGAFGPDGVNSDD--SPAPLSHILQ 554
 ++ + L + E +T L D + ++ S LS+IL+
 352 RKQEILSRKKKSQPSKYTNSEGEIVTDLRTLHTTLSAPAESDSITEEEKSSACTLSYILE 411

 553 QLPALHRPSSLQORDNYKRTPLHYSAQYGLCEXTRIILQALSEWDADNADVAIDDDIDVWGD 374
 +LP HLRP L Q DNYKRTPLHYS QYGL E T++I++ + EW+ WN ++ IDD+ +GD
 412 ELPIHLRPCLFQHDNYKRTPLHYSCQYGLSEVTCLIILKLMKEWNIWN-EIPIDDVSAFGD 470

 373 SENLTPLHLAVIGTHPLTVSTLLSFMNPEKSLNSPRLLHLATRXXXXXXXXXXXXXSAKGFD 194
 +E+LTPLHL V+G HP T LL ++P L S LLHLAT S+K FD
 471 AESLTPLHLCVLGAHPKTTEVLLQSLDPNVKLKSSSLLHLATEWNNYPILLHVLLSSKRFD 530

 193 IDYQEPENLETALYVACKLDIYEAAEYLVKQGANMELGEKLFGWTPIFAAAATEGYARIVQ 14
 I+YQ+ E ET LY+AC+L+ +EAA L+ GA++E+ EKLFGWT IF AA EG+ IV+
 531 INYQDNELHETPLYLACRLNFFEAACVCLLYNGADLEIREKLFGWTAIFVAAAEGFTDIVK 590

 13 LLV 5 A. gossypii
 LL+
 591 LLI 593 S. cerevisiae

Fig. 10**Oligo 167**

931 DASESANSWVWCAVMTLFLVGGTCFWQVNHLRTFFMKQKVL 806

DASE+ NS +W V+ L VL TC WQ+ HL FF+KQK+L

178 DASEAVNSRAMWWIVIQLIVLAVTCGWQMKHLGKFFVKQKIL 219

*A. gossypii**S. cerevisiae*

SEQUENCE LISTING

<110> BASF Aktiengesellschaft

<120> Neue Polynukleotide zur Verbesserung des Transmembran-Transports

<130> M/42291_PCT

<160> 45

<170> PatentIn version 3.1

<210> 1

<211> 1253

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 19

<220>

<221> misc_feature

<222> (579) .. (579)

<223> unknown nucleotide

<400> 1
gatctccaag cctttaaata ggcacaggta acccataacc gaaacctcat gccgcatctc 60
tcgcagcagg tcgcctatca gatatatgtg acgctctgtg tctcttgccg ggggcacact 120
ctggagccgt gtgcgcaata gctccagcgg cgcaatagta gttgcagcca atatccgcgc 180
aaatgctcca cacactagtg gggttgccac aggtagccgg gatgccaagg gcgagttatc 240

acgcagtgct tcatagccgg aaaagtacac cacgttagcc ggcaccgcca ttacgagcgt 300
aatccccagc cctcgccaca gaggcggcag accctctagt tgggctatct tcctgagccc 360
ctccagagtt ccctgcagcc tcgcagcagg ctctcggcag cgcacgtttg caaagcactc 420
atcctgccaa aacactttcc ctgccggttt cgacagctgt cctgtacatg tacaacttgg 480
aagcattttct tgctgttgaa gccgcacgcg gaccacgtcc agtggcgtga gaaacagtga 540
cgtgaccagc gagccagcag aggcgcttac catgcgttnc ttcagcgtaa ctttctttcc 600
gggttcctta ctcatggccc caaatttgca cctttcagcg ctaaccccg tccaatagtag 660
tgccctcttc tgtttagccta tcttgccgct cgagagtaac ttatgttttg atcgaagctt 720
tcttgggggc gattattcgc tgtgtatggg acttagtttt actatgcgct aaaatgttgt 780
cggggcccca aagccgttgt acggcgctcag cgttgatgtg caccgcgtgg ctatttaata 840
atatactata taaatactaa aactattaag tattaatcat gatgatgtgg aatcagtatt 900
tcagtaagat gtagcatcag agttgtccaa tgggacctcc gggctttcct cgtatagtct 960
taaaaccatg gaggttatcc gtttcagctg gaattgtaat agttcgtcga cggttgccat 1020
attaatctgg gtcttctccc cgcgctggaa caaagacatg tcaaattcat cctcctcca 1080
gttcattgtac tcgtcaatgc tatagcagtc gtgagtcagg gtactcttcg gtgaacaata 1140
tgtattggaa cattttctat ccattgggta acctctctgc ggtggggcat gttcagcgct 1200
gcgcgcggat taaaataggc aggaaagctc gactgtggaa catcgtaggg atc 1253

<210> 2

<211> 203

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 19

<220>

<221> misc_feature

<222> (11)..(11)

<223> unknown amino acid

<400> 2

Lys Glu Pro Gly Lys Lys Val Thr Leu Lys Xaa Arg Met Val Ser Ala
 1 5 10 15

Ser Ala Gly Ser Leu Val Thr Ser Leu Phe Leu Thr Pro Leu Asp Val
 20 25 30

Val Arg Val Arg Leu Gln Gln Gln Glu Met Leu Pro Ser Cys Thr Cys
 35 40 45

Thr Gly Gln Leu Ser Lys Pro Ala Gly Lys Val Phe Trp Gln Asp Glu
 50 55 60

Cys Phe Ala Asn Val Gly Cys Arg Glu Pro Ala Ala Arg Leu Gln Gly
 65 70 75 80

Thr Leu Glu Gly Leu Arg Lys Ile Ala Gln Leu Glu Gly Leu Pro Thr
 85 90 95

Leu Trp Arg Gly Leu Gly Ile Thr Leu Val Met Ala Val Pro Ala Asn
 100 105 110

Val Val Tyr Phe Ser Gly Tyr Glu Ala Leu Arg Asp Asn Ser Pro Leu
 115 120 125

Ala Ser Arg Leu Pro Val Ala Asn Pro Leu Val Cys Gly Ala Phe Ala
 130 135 140

Arg Ile Leu Ala Ala Thr Thr Ile Ala Pro Leu Glu Leu Leu Arg Thr
 145 150 155 160

Arg Leu Gln Ser Val Pro Arg Ala Arg Asp Thr Glu Arg Thr Ile Tyr
 165 170 175

Leu Ile Gly Asp Leu Leu Arg Glu Met Arg His Glu Val Ser Val Met
 180 185 190

Gly Tyr Arg Ala Leu Phe Lys Gly Leu Glu Ile
 195 200

<210> 3

<211> 2020

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (112) .. (294)

<223>

<220>

<221> misc_feature

<223> Oligo 19

<400> 3

```

aatcccccta tattcaacaa cagtccgcag tatttcccag atctgaacta tttatctcca      60
tggcacagtt attcaaacgg tgtcggcacc ccaataccaa ttaacagttc c atg att      117
                                     Met Ile
                                     1

ggg aac cca aac ctt tcc aac gtc ccc atg acc aag acg tac gat ccc      165
Gly Asn Pro Asn Leu Ser Asn Val Pro Met Thr Lys Thr Tyr Asp Pro
      5                      10                      15

tac gat gtt cca cag tcg agc ttt cct gcc tat ttt aat ccg cgc gca      213
Tyr Asp Val Pro Gln Ser Ser Phe Pro Ala Tyr Phe Asn Pro Arg Ala
      20                      25                      30

gcg ctg aac atg ccc cac cgc gag aag gtt aac caa tgg ata gaa aat      261
Ala Leu Asn Met Pro His Arg Glu Lys Val Asn Gln Trp Ile Glu Asn
      35                      40                      45                      50

gtt cca ata cat att gtt cac cga aga gta ccc tgactcacga ctgctatagc      314
Val Pro Ile His Ile Val His Arg Arg Val Pro
      55                      60

attgacgagt acatgaactg ggaggaggat gaatttgaca tgtctttgtt ccagcgcggg      374

gagaagaccc agattaatat ggcaaccgtc gacgaactat tacaattcca gctgaaacgg      434

ataacctcca tggttttaag actatacgag gaaagcccgg aggtcccatt ggacaactct      494

gatgctacat cttactgaaa tactgattcc acatcatcat gattaatact taatagtttt      554

agtattttata tagtatatta ttaaatagcc acgggggtgca catcaacgct gacgccgtac      614

aacggcttttc gggccccgac aacatttttag cgcatagtaa aactaagtcc catacacagc      674

gaataatcgc cccaagaaa gcttcgatca aaacataagt tactctcgag cggcaagata      734

ggctaacagg aagaggcagt actattggac gggggttagcg ctgaaagggtg caaatttggg      794

```

```

gccatgagta aggaacccgg aaagaagggt acgctgaagg aacgcatggt aagcgccctct      854
gctggctcgc tgggtcacgtc actgtttctc acgccactgg acgtgggtccg cgtgcggtct      914
caacagcaag aaatgcttcc aagttgtaca tgtacaggac agctgtcgaa accggcaggg      974
aaagtgtttt ggcaggatga gtgctttgca aacgtcggct gccgagagcc tgctgcgagg     1034
ctgcagggaa ctctggaggg gctcaggaag atagcccaac tagaggggtct gccgactctg     1094
tggtcgagggc tggggattac gctcgtaatg gcggtgccgg ctaacgtggt gtacttttcc     1154
ggctatgaag cactgcgtga taactcgccc ttggcatccc ggctacctgt ggcaaaccce      1214
ctagtgtgtg gagcatttgc gcggatattg gctgcaacta ctattgcgcc gctggagcta     1274
ttgcgcacac ggctccagag tgtgccccgc gcaagagaca cagagcgtag aatatatctg     1334
ataggcgacc tgctgcgaga gatgcggcat gaggtttcgg ttatgggtta ccgtgcgcta     1394
tttaaaggct tggagatcac tttatggagg gacgtgccct tcagcgcaat ctattgggga     1454
acatacgagt tctgtaaaac ccagttctgg gcccgccatg ccgcaaccce taatgcatca     1514
aactgggacc atttcatcgg cagttttgcc tgcggtagca tgggcgggtgc tgttgcagca     1574
cttttgacac atccttttga tgtgggcaag acccgcatgc agattgcgat tgccagtcca     1634
cagcagctaa ctgtggggggg aaaagctacg aaaactgatg actcaagagg catgttctca     1694
tttttgaatg ccattaggaa atcagaaggt attagagcgc tatataccgg cctattacct     1754
agggtgatga agattgcacc aagttgcgcc ataatgatct cgacttatga actgtcgaag     1814
aagttcttca ctagtgaac tttgtttctt cctgtatata ccaagtaatt cacactgttg     1874
atacacgata cagcattttt cttaaaaccc tgtgatatac gcttgctgta gtacacgccc     1934
aaacggtgag gtctataatt ttacatgtcc gcgagatggg attggatcgg cagaatttta     1994
gagctgccgc tgtagcccat ccgtcg                                     2020

```

<210> 4

<211> 61

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 19

<400> 4

Met Ile Gly Asn Pro Asn Leu Ser Asn Val Pro Met Thr Lys Thr Tyr
 1 5 10 15

Asp Pro Tyr Asp Val Pro Gln Ser Ser Phe Pro Ala Tyr Phe Asn Pro
 20 25 30

Arg Ala Ala Leu Asn Met Pro His Arg Glu Lys Val Asn Gln Trp Ile
 35 40 45

Glu Asn Val Pro Ile His Ile Val His Arg Arg Val Pro
 50 55 60

<210> 5

<211> 1495

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 24

<220>

<221> misc_feature

<222> (1386)..(1386)

<223> unknown nucleotide

<400> 5

gacgccacc tgggccctgt tcagcgcgcc ggacgtctgc tgcgcctcca gcggtccag	60
gaacttcagc atcgcggggt cgatctgcga aaagcacttg tacagcgggt agaagaggcc	120
acgcccgtcc tggatccagt cgaagtactg tggcatgtgc gacaccacca cgtccctaga	180
aaacgtcccc ttcagcgaca cttcagatc ctgcagcagg ttgaccaatc gcaacaggtc	240
ctgcttgatc tcctccgtgc tcaggttctt gtttccgatt cgcagcatca catcctccag	300
tgagtcagct ccgggcccta gtgtggtgc taaggggggc gagataaacg cgtaggtac	360
ctgtatcatg ctgcggacac cacctcactt tgaaatatta caactgtgct atcagctgcc	420
ctttataagc agcgtgtct ccatacaatt gcgctgattt tatgtagtgt atgcgtgtaa	480

```

tctgcttttt ccctattgag acgtcagcat ctcgtagcaa ggctccgtgt gcttacttaa      540
gaaaagatgg aatcgatgga atcgatgaaa ctaaaacatt ttatcaaata tcagttctat      600
atatatgtag ttattgggag ataatacggg tgaatctggg gcatgtctta aggcatggg      660
gcacggggccg tgtcctccag cacatgcttt atgacctctt catttaggcg aagatttccg      720
ccctcgatat cgcgggggaat gtctgcggcc tgtggagcgg aagcctcttc ctgggcagga      780
gctggagacg cctcgggctt cgcaggctcc gacgcgtgtt cgttcaatag cttgaacaac      840
tcgctgtcct cgtcacggta taggtcggca aacttgccgg cctcgatgac ggagccgtca      900
ttgccgagaa ccacgatgtt ctcggagcgg cgtatggtgc tgaggcgggtg ggcgatgctc      960
acgatggtga tctccttgct gcgcattagc ctgccgaggg tgtagttgat ggcgccctcg     1020
ctttcaacgt ccaacgccga ggtggcctcg tcgaggatta ggattttcgg tttcttgatg     1080
agcgcgcggg cgatggcaat gcgctggcgc tggccccgcg tgagtagggc gcctctaggc     1140
ccaatcacgg tgtcatagcc gtccgggaac ttggtgatga agttatggca gaagcacttc     1200
ttggcaactg cgcgtatctc atccatggag ggcgtgtagg aaaggccgta ggtgatgttg     1260
tctctgatgg tcccggacat caagaccggc tcctgctgga caacgccaag atgttttctg     1320
agcgacttgg cgcttacctt ggatatgtcc tgaccgtcta tgaggatggt gccgctgttg     1380
gcatcngtaa aaccgtagga ggagggacgc aatggtggac ttgccccgcc cggagggccc     1440
gacgatgcag acgttggagc ccgggggtat cgtgaggcgg agtttcttca agatc         1495

```

<210> 6

<211> 36

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 24

<400> 6

```

Ile Leu Lys Lys Leu Arg Leu Thr Ile Thr Pro Gly Ser Asn Val Cys
1           5           10          15

```

```

Ile Val Gly Pro Ser Gly Arg Gly Lys Ser Thr Ile Ala Ser Leu Leu
20          25          30

```

Leu Arg Phe Tyr
35

<210> 7

<211> 225

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 24

<400> 7

Gly Pro Pro Gly Gly Ala Ser Pro Pro Leu Arg Pro Ser Ser Tyr Gly
1 5 10 15

Phe Thr Asp Ala Asn Ser Gly Thr Ile Leu Ile Asp Gly Gln Asp Ile
20 25 30

Ser Lys Val Ser Ala Lys Ser Leu Arg Lys His Leu Gly Val Val Gln
35 40 45

Gln Glu Pro Val Leu Met Ser Gly Thr Ile Arg Asp Asn Ile Thr Tyr
50 55 60

Gly Leu Ser Tyr Thr Pro Ser Met Asp Glu Ile Arg Ala Val Ala Lys
65 70 75 80

Lys Cys Phe Cys His Asn Phe Ile Thr Lys Phe Pro Asp Gly Tyr Asp
85 90 95

Thr Val Ile Gly Pro Arg Gly Ala Leu Leu Ser Gly Gly Gln Arg Gln
100 105 110

Arg Ile Ala Ile Ala Arg Ala Leu Ile Lys Lys Pro Lys Ile Leu Ile
115 120 125

Leu Asp Glu Ala Thr Ser Ala Leu Asp Val Glu Ser Glu Gly Ala Ile
130 135 140

Asn Tyr Thr Leu Gly Arg Leu Met Arg Ser Lys Glu Ile Thr Ile Val

145		150		155		160
Ser Ile Ala His	Arg Leu Ser Thr Ile	Arg Arg Ser Glu Asn Ile Val				
	165	170			175	
Val Leu Gly Asn Asp Gly Ser Val Ile Glu Ala Gly Lys Phe Ala Asp						
	180	185			190	
Leu Tyr Arg Asp Glu Asp Ser Glu Leu Phe Lys Leu Leu Asn Glu His						
	195	200			205	
Ala Ser Glu Pro Ala Lys Pro Glu Ala Ser Pro Ala Pro Ala Gln Glu						
	210	215			220	

Glu
225

<210> 8
 <211> 4118
 <212> DNA
 <213> Ashbya gossypii

<220>
 <221> CDS
 <222> (820)..(3081)
 <223>

<220>
 <221> misc_feature
 <223> Oligo 24

<400> 8
 ctgtcacagc aagacctggc ttcccgtcac gctatggctc gcctgatcat ctcatctcgc 60
 ctgcgcatcg agcgcagctc aaaattttta ccattacccc ggccagtaca catattacat 120
 aatcaccagt gtccagggat taccgggaca tcaaacaatca cagagaacga acccaatgaa 180
 atgcaggaaa agctgctcag cgccggtgcc cggcggccat taagcgtgca cctacagcca 240
 cttcactcca ggccacacac tccacgcctc cccgcaagta aacacgatgt ctgcctggag 300

aaaagccggt ttgacctaca atagctacct ggccgttgcg gccaggaccg tgcgcgcggc	360
tctgaagaag gagctccaga gccctgccgt cctcaaccga tctgtgacgg aggcaaagg	420
catcgactac gcctccaagg gctccgccgc ggaggccgtg ccgctacgga agtgagcagc	480
acgcgtgctc ttctcgttga actgctggca cgcagcctgg gccctgttg	540
gcggatggag acgcccagca caggcccttt gacgggctgt gctggtggca tgtgtgtagt	600
tcgactatat aaactctcgt tctctaata taatgacaga aggtcgattg ggcgatttcg	660
gtgctgcgaa cggggacggt gtgcaggggc aaaggcgggt ttcgcgggca agcgaagagg	720
tgcgttttcc gcaagcttgc ttgtgtcccc gcccaaggca agacacagtc tactcgcatc	780
agttttggca gagcgtgttg ctacaggatt ccaggcgcgt atg tgg aga gcg ata	834
	Met Trp Arg Ala Ile
	1 5
agg ccc ttt ggg ggg ctt ggc cct gta gcg gtg tta cag agg cta ccg	882
Arg Pro Phe Gly Gly Leu Gly Pro Val Ala Val Leu Gln Arg Leu Pro	
	10 15 20
ggc ccg cgg cac aat gtc gcc gtg agg ctt cag gtg atg tcc agg gcc	930
Gly Pro Arg His Asn Val Ala Val Arg Leu Gln Val Met Ser Arg Ala	
	25 30 35
aga tgc ggc tgg cag tgt gtg cgt gcg aca tct cga ctg aca gtg ccg	978
Arg Cys Gly Trp Gln Cys Val Arg Ala Thr Ser Arg Leu Thr Val Pro	
	40 45 50
gtg tgc agg caa agc cgg gcg tac agc gct gcc cgt gtg ggc ggc gag	1026
Val Cys Arg Gln Ser Arg Ala Tyr Ser Ala Ala Arg Val Gly Gly Glu	
	55 60 65
gga acg cct tcg ccg cgg gac aag gag cat gag cag gag cgc aaa gtt	1074
Gly Thr Pro Ser Pro Arg Asp Lys Glu His Glu Gln Glu Arg Lys Val	
	70 75 80 85
gaa aag cga ccg gcg gag ccg tcg cgc gcg caa ccc gca aag cca gcc	1122
Glu Lys Arg Pro Ala Glu Pro Ser Arg Ala Gln Pro Ala Lys Pro Ala	
	90 95 100
ggc tgg cac gag gtg ttc cgg ctc ttg cga ttg gtg aag ccc gac tgg	1170
Gly Trp His Glu Val Phe Arg Leu Leu Arg Leu Val Lys Pro Asp Trp	
	105 110 115
aag ttg ttg ctt gcg gcg ttg ggg ctc ctc acg ata tct tgc tcg att	1218
Lys Leu Leu Leu Ala Ala Leu Gly Leu Leu Thr Ile Ser Cys Ser Ile	
	120 125 130
ggg atg gcg gtt ccg aag gtt att ggg ctt gtg cta gat gcc acg aaa	1266
Gly Met Ala Val Pro Lys Val Ile Gly Leu Val Leu Asp Ala Thr Lys	
	135 140 145
gac gca gtc gca cag gca gaa aag aac gaa acg ggc gcg ccg atg tct	1314
Asp Ala Val Ala Gln Ala Glu Lys Asn Glu Thr Gly Ala Pro Met Ser	
	150 155 160 165
ttg cta gac atg ccg cct atc atc ggg aac ttc acg cta gtc gac gtg	1362
Leu Leu Asp Met Pro Pro Ile Ile Gly Asn Phe Thr Leu Val Asp Val	

170	175	180	
ctt gct gcg ttt gcg ggc gcg ttg ctc gtg ggc tcc gcc gcg aac ttc Leu Ala Ala Phe Ala Gly Ala Leu Leu Val Gly Ser Ala Ala Asn Phe 185 190 195			1410
ggg cgg atg ttc ctc ttg aaa atg ctt ggt gag cgt ttg gtt gcg cgg Gly Arg Met Phe Leu Leu Lys Met Leu Gly Glu Arg Leu Val Ala Arg 200 205 210			1458
cta cgt gcg cag gtc atc aag aag acg ttg cat cac gac gcg gag ttt Leu Arg Ala Gln Val Ile Lys Lys Thr Leu His His Asp Ala Glu Phe 215 220 225			1506
ttc gac agg aac aag gtc ggc gac ttg ata tcg cgg ctt ggt tcc gac Phe Asp Arg Asn Lys Val Gly Asp Leu Ile Ser Arg Leu Gly Ser Asp 230 235 240 245			1554
gcc tac att gta tcg cgc tcg atg acg cag aac att tcg gac ggg gtt Ala Tyr Ile Val Ser Arg Ser Met Thr Gln Asn Ile Ser Asp Gly Val 250 255 260			1602
aaa gcc gcg ctc tgc gga ggc gtc ggt gtg ggc atg atg ttc cac att Lys Ala Ala Leu Cys Gly Gly Val Gly Val Gly Met Met Phe His Ile 265 270 275			1650
tcg agc aca ttg act tct gca atg ata atg ttt gcc cct ccc ctc ctc Ser Ser Thr Leu Thr Ser Ala Met Ile Met Phe Ala Pro Pro Leu Leu 280 285 290			1698
atc gcc gcc acc atc tac ggc aag cgc att cgc gcc atc tcg cgc gag Ile Ala Ala Thr Ile Tyr Gly Lys Arg Ile Arg Ala Ile Ser Arg Glu 295 300 305			1746
ctc cag caa tct aca ggt aac ttg acg cgg gtg tca gag gag cag ttc Leu Gln Gln Ser Thr Gly Asn Leu Thr Arg Val Ser Glu Glu Gln Phe 310 315 320 325			1794
aat ggc atc aag acc atc aag tca ttc gtc gcg gag gga aag gag atg Asn Gly Ile Lys Thr Ile Lys Ser Phe Val Ala Glu Gly Lys Glu Met 330 335 340			1842
cgg aga tac aac acg gcc gtg cgg agg ctc ttc aac gtt gca aag agg Arg Arg Tyr Asn Thr Ala Val Arg Arg Leu Phe Asn Val Ala Lys Arg 345 350 355			1890
gaa ggt ata acg agc gcg acg ttt ttc agc ggt acg aat gtg ttg ggc Glu Gly Ile Thr Ser Ala Thr Phe Phe Ser Gly Thr Asn Val Leu Gly 360 365 370			1938
gac gtg agt ttt ttg ctt gtt ttg gcg tat ggg tcg cac ctc gtt ttg Asp Val Ser Phe Leu Leu Val Leu Ala Tyr Gly Ser His Leu Val Leu 375 380 385			1986
cag ggc gaa cta tcc ctt ggg agc ttg acg gcg ttc atg ctc tac act Gln Gly Glu Leu Ser Leu Gly Ser Leu Thr Ala Phe Met Leu Tyr Thr 390 395 400 405			2034
gag tac act gga agc tcg gtt ttc ggt tta tcg aac ttt tac tca gag Glu Tyr Thr Gly Ser Ser Val Phe Gly Leu Ser Asn Phe Tyr Ser Glu 410 415 420			2082

ttg atg aag ggc gct ggt gct gct tcg cgc ttg ttc gaa ttg aca gac	2130
Leu Met Lys Gly Ala Gly Ala Ala Ser Arg Leu Phe Glu Leu Thr Asp	
425 430 435	
tac aag aac tcc gtg ccg tct acc gtt ggc cag acg ttt gtg ccg cgc	2178
Tyr Lys Asn Ser Val Pro Ser Thr Val Gly Gln Thr Phe Val Pro Arg	
440 445 450	
gat ggt cgc atc gag ttc aag gac gtg tcc ttc agc tac ccc acg cgg	2226
Asp Gly Arg Ile Glu Phe Lys Asp Val Ser Phe Ser Tyr Pro Thr Arg	
455 460 465	
ccg acg agc cag atc ttc aag aaa ctc cgc ctc acg ata acc ccg ggc	2274
Pro Thr Ser Gln Ile Phe Lys Lys Leu Arg Leu Thr Ile Thr Pro Gly	
470 475 480 485	
tcc aac gtc tgc atc gtc ggg ccc tcc ggg cgg ggc aag tcc acc att	2322
Ser Asn Val Cys Ile Val Gly Pro Ser Gly Arg Gly Lys Ser Thr Ile	
490 495 500	
gcg tcc ctc ctc cta cgg ttt tac gat gcc aac agc ggc acc atc ctc	2370
Ala Ser Leu Leu Leu Arg Phe Tyr Asp Ala Asn Ser Gly Thr Ile Leu	
505 510 515	
ata gac ggt cag gac ata tcc aag gta agc gcc aag tcg ctc aga aaa	2418
Ile Asp Gly Gln Asp Ile Ser Lys Val Ser Ala Lys Ser Leu Arg Lys	
520 525 530	
cat ctt ggc gtt gtc cag cag gag ccg gtc ttg atg tcc ggg acc atc	2466
His Leu Gly Val Val Gln Gln Glu Pro Val Leu Met Ser Gly Thr Ile	
535 540 545	
aga gac aac atc acc tac ggc ctt tcc tac acg ccc tcc atg gat gag	2514
Arg Asp Asn Ile Thr Tyr Gly Leu Ser Tyr Thr Pro Ser Met Asp Glu	
550 555 560 565	
ata cgc gca gtt gcc aag aag tgc ttc tgc cat aac ttc atc acc aag	2562
Ile Arg Ala Val Ala Lys Lys Cys Phe Cys His Asn Phe Ile Thr Lys	
570 575 580	
ttc ccg gac ggc tat gac acc gtg att ggg cct aga ggc gcc cta ctc	2610
Phe Pro Asp Gly Tyr Asp Thr Val Ile Gly Pro Arg Gly Ala Leu Leu	
585 590 595	
agc ggg ggc cag cgc cag cgc att gcc atc gcc cgc gcg ctc atc aag	2658
Ser Gly Gly Gln Arg Gln Arg Ile Ala Ile Ala Arg Ala Leu Ile Lys	
600 605 610	
aaa ccg aaa atc cta atc ctc gac gag gcc acc tcg gcg ttg gac gtt	2706
Lys Pro Lys Ile Leu Ile Leu Asp Glu Ala Thr Ser Ala Leu Asp Val	
615 620 625	
gaa agc gag ggc gcc atc aac tac acc ctc ggc agg cta atg cgc agc	2754
Glu Ser Glu Gly Ala Ile Asn Tyr Thr Leu Gly Arg Leu Met Arg Ser	
630 635 640 645	
aag gag atc acc atc gtg agc atc gcc cac cgc ctc agc acc ata cgc	2802
Lys Glu Ile Thr Ile Val Ser Ile Ala His Arg Leu Ser Thr Ile Arg	
650 655 660	
cgc tcc gag aac atc gtg gtt ctc ggc aat gac ggc tcc gtc atc gag	2850
Arg Ser Glu Asn Ile Val Val Leu Gly Asn Asp Gly Ser Val Ile Glu	

665	670	675	
gcc ggc aag ttt gcc gac cta tac cgt gac gag gac agc gag ttg ttc Ala Gly Lys Phe Ala Asp Leu Tyr Arg Asp Glu Asp Ser Glu Leu Phe 680 685 690			2898
aag cta ttg aac gaa cac gcg tcg gag cct gcg aag ccc gag gcg tct Lys Leu Leu Asn Glu His Ala Ser Glu Pro Ala Lys Pro Glu Ala Ser 695 700 705			2946
cca gct cct gcc cag gaa gag gct tcc gct cca cag gcc gca gac att Pro Ala Pro Ala Gln Glu Glu Ala Ser Ala Pro Gln Ala Ala Asp Ile 710 715 720 725			2994
ccc cgc gat atc gag ggc gga aat ctt cgc cta aat gaa gag gtc ata Pro Arg Asp Ile Glu Gly Gly Asn Leu Arg Leu Asn Glu Glu Val Ile 730 735 740			3042
aag cat gtg ctg gag gac acg gcc cgt gcc cca atg cct taagacatgc Lys His Val Leu Glu Asp Thr Ala Arg Ala Pro Met Pro 745 750			3091
cccagattca cccgtattat ctccaataa ctacatatat atagaactga tatttgataa			3151
aatgttttag tttcatcgat tccatcgatt ccatcttttc ttaagtaagc acacggagcc			3211
ttgctacgag atgctgacgt ctcaataggg aaaaagcaga ttacacgcat caactacata			3271
aaatcagcgc aattgtatgg agacagcgct gcttataaag ggcagctgat agcacagttg			3331
taatatttca aagtgaggtg gtgtccgcag catgatacag gtacctaacg cgtttatctc			3391
gccccctta gcgaccacac tagggcccg agctgactca ctggaggatg tgatgctgcg			3451
aatcggaaac aagaacctga gcacggagga gatcaagcag gacctgttgc gattggtcaa			3511
cctgctgcag gatctgaagg tgctcgctgaa ggggacgttt tctagggacg tggtggtgtc			3571
gcacatgcc aagtacttcg actggatcca ggacgggcgt ggcctcttct acccgctgta			3631
caagtgcctt tcgcagatcg agcccgcat gctgaagttc ctggaccgc tggaggcgca			3691
gcagacgtcc ggcgcgtga acagggacca ggtggcgatc ttcaacctga tcgaccaggt			3751
gtccgagctc gtgctgagcc tgaagccgct gttccagtcg gtgaagaacc tattcgacac			3811
ggcgctggag ttcaatgaga ttttcaagga ccacatgaac tcgctactgg aggagatcga			3871
aggaaacatg aagaagtgcc tcgcgctgca ccaggactgc ttcgcgtcgc ccgtgagaca			3931
ccctccgtcc ttcacgttgg accagctcgt ggagcttata tcctcatcgt caaacaacca			3991
gcggctgcag atgccgacct tcaatccctt ggaaaagcgc atctaccagg gattactgcg			4051
agctaagaga atgctattgt gccgatccag acatcgctga agagacgtcc tgaaaacccg			4111
catccca			4118

<211> 754

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 24

<400> 9

Met Trp Arg Ala Ile Arg Pro Phe Gly Gly Leu Gly Pro Val Ala Val
 1 5 10 15

Leu Gln Arg Leu Pro Gly Pro Arg His Asn Val Ala Val Arg Leu Gln
 20 25 30

Val Met Ser Arg Ala Arg Cys Gly Trp Gln Cys Val Arg Ala Thr Ser
 35 40 45

Arg Leu Thr Val Pro Val Cys Arg Gln Ser Arg Ala Tyr Ser Ala Ala
 50 55 60

Arg Val Gly Gly Glu Gly Thr Pro Ser Pro Arg Asp Lys Glu His Glu
 65 70 75 80

Gln Glu Arg Lys Val Glu Lys Arg Pro Ala Glu Pro Ser Arg Ala Gln
 85 90 95

Pro Ala Lys Pro Ala Gly Trp His Glu Val Phe Arg Leu Leu Arg Leu
 100 105 110

Val Lys Pro Asp Trp Lys Leu Leu Leu Ala Ala Leu Gly Leu Leu Thr
 115 120 125

Ile Ser Cys Ser Ile Gly Met Ala Val Pro Lys Val Ile Gly Leu Val
 130 135 140

Leu Asp Ala Thr Lys Asp Ala Val Ala Gln Ala Glu Lys Asn Glu Thr
 145 150 155 160

Gly Ala Pro Met Ser Leu Leu Asp Met Pro Pro Ile Ile Gly Asn Phe
 165 170 175

Thr Leu Val Asp Val Leu Ala Ala Phe Ala Gly Ala Leu Leu Val Gly
 180 185 190

Ser Ala Ala Asn Phe Gly Arg Met Phe Leu Leu Lys Met Leu Gly Glu
 195 200 205

Arg Leu Val Ala Arg Leu Arg Ala Gln Val Ile Lys Lys Thr Leu His
 210 215 220

His Asp Ala Glu Phe Phe Asp Arg Asn Lys Val Gly Asp Leu Ile Ser
 225 230 235 240

Arg Leu Gly Ser Asp Ala Tyr Ile Val Ser Arg Ser Met Thr Gln Asn
 245 250 255

Ile Ser Asp Gly Val Lys Ala Ala Leu Cys Gly Gly Val Gly Val Gly
 260 265 270

Met Met Phe His Ile Ser Ser Thr Leu Thr Ser Ala Met Ile Met Phe
 275 280 285

Ala Pro Pro Leu Leu Ile Ala Ala Thr Ile Tyr Gly Lys Arg Ile Arg
 290 295 300

Ala Ile Ser Arg Glu Leu Gln Gln Ser Thr Gly Asn Leu Thr Arg Val
 305 310 315 320

Ser Glu Glu Gln Phe Asn Gly Ile Lys Thr Ile Lys Ser Phe Val Ala
 325 330 335

Glu Gly Lys Glu Met Arg Arg Tyr Asn Thr Ala Val Arg Arg Leu Phe
 340 345 350

Asn Val Ala Lys Arg Glu Gly Ile Thr Ser Ala Thr Phe Phe Ser Gly
 355 360 365

Thr Asn Val Leu Gly Asp Val Ser Phe Leu Leu Val Leu Ala Tyr Gly
 370 375 380

Ser His Leu Val Leu Gln Gly Glu Leu Ser Leu Gly Ser Leu Thr Ala
 385 390 395 400

Phe Met Leu Tyr Thr Glu Tyr Thr Gly Ser Ser Val Phe Gly Leu Ser
 405 410 415

Asn Phe Tyr Ser Glu Leu Met Lys Gly Ala Gly Ala Ala Ser Arg Leu
 420 425 430

Phe Glu Leu Thr Asp Tyr Lys Asn Ser Val Pro Ser Thr Val Gly Gln
 435 440 445

Thr Phe Val Pro Arg Asp Gly Arg Ile Glu Phe Lys Asp Val Ser Phe
 450 455 460

Ser Tyr Pro Thr Arg Pro Thr Ser Gln Ile Phe Lys Lys Leu Arg Leu
 465 470 475 480

Thr Ile Thr Pro Gly Ser Asn Val Cys Ile Val Gly Pro Ser Gly Arg
 485 490 495

Gly Lys Ser Thr Ile Ala Ser Leu Leu Leu Arg Phe Tyr Asp Ala Asn
 500 505 510

Ser Gly Thr Ile Leu Ile Asp Gly Gln Asp Ile Ser Lys Val Ser Ala
 515 520 525

Lys Ser Leu Arg Lys His Leu Gly Val Val Gln Gln Glu Pro Val Leu
 530 535 540

Met Ser Gly Thr Ile Arg Asp Asn Ile Thr Tyr Gly Leu Ser Tyr Thr
 545 550 555 560

Pro Ser Met Asp Glu Ile Arg Ala Val Ala Lys Lys Cys Phe Cys His
 565 570 575

Asn Phe Ile Thr Lys Phe Pro Asp Gly Tyr Asp Thr Val Ile Gly Pro
 580 585 590

Arg Gly Ala Leu Leu Ser Gly Gly Gln Arg Gln Arg Ile Ala Ile Ala
 595 600 605

Arg Ala Leu Ile Lys Lys Pro Lys Ile Leu Ile Leu Asp Glu Ala Thr
 610 615 620

Ser Ala Leu Asp Val Glu Ser Glu Gly Ala Ile Asn Tyr Thr Leu Gly
 625 630 635 640

Arg Leu Met Arg Ser Lys Glu Ile Thr Ile Val Ser Ile Ala His Arg
 645 650 655

Leu Ser Thr Ile Arg Arg Ser Glu Asn Ile Val Val Leu Gly Asn Asp
 660 665 670

Gly Ser Val Ile Glu Ala Gly Lys Phe Ala Asp Leu Tyr Arg Asp Glu
 675 680 685

Asp Ser Glu Leu Phe Lys Leu Leu Asn Glu His Ala Ser Glu Pro Ala
 690 695 700

Lys Pro Glu Ala Ser Pro Ala Pro Ala Gln Glu Glu Ala Ser Ala Pro
 705 710 715 720

Gln Ala Ala Asp Ile Pro Arg Asp Ile Glu Gly Gly Asn Leu Arg Leu
 725 730 735

Asn Glu Glu Val Ile Lys His Val Leu Glu Asp Thr Ala Arg Ala Pro
 740 745 750

Met Pro

<210> 10

<211> 1002

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 109

<400> 10
 gatcggagcc gattaaaacg gcacaagctt gggagctgat cgaactactt tcacaaaatg 60
 atgtcgtgaa gtatggagat atcgttttca ggccgttggt taagtcttct ccagaggccg 120
 ggttgttgga gttagagaaa aatgggctga tcacaatcag ccgcaataga ggagtgttgc 180
 aagatatccg gcccgctaaa ccattgttta aagccgcatt cagctacctg ttacaggaca 240
 aggacctgtc catcgtcctc cgaacagggt actacttgcg actaatcgcg tttgaaactg 300
 ggcggatcaa aaagtgggag gaggaattgc gtctattggc aaagggttagt gaccaaagaa 360
 tatgcaagag tcggttgaat tacttagcaa gcaaaataga cgccagtagt ggggtaatta 420
 acagttgtga ggacaaggta aaagaaatgt caaagcgtat ttagaaacgc tcgcatcaat 480
 aagaccccat ttgccaoga tctggcaaga tgaactaaat agcgataatc acactaggtc 540
 tcattagttt agcattccag ctgcacgggc tttagtttag tacctcaatt ccaaaaacgg 600
 cactatcttg tccgagcgct ggtgaatatc cttgccgtat gtgaaattat ccagggttga 660

```

tgtagttggc cggagctacc agtaattgtt ggcctatgca cgggccgcta ttttccgtaa    720
cataatccag agatccgtta acaggcaaca tgaactgatt gattaatttt aatgaacttt    780
tttttttggc tttgaaaggc ctaaaattgg atcatctaca ttccctgacg atataggatg    840
gtaaaagccc taccaaccaa gaaattctga tccctttgac aaatccctta taacatagat    900
attagaatat aatcaacccc attgtgtgac aactcttctc caaaaataat tgggtctttca    960
cacaattaaa gattggcttc aaggtagctc cacggtgaga tc                        1002

```

<210> 11

<211> 147

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 109

<400> 11

```

Lys Thr Ala Gln Ala Trp Glu Leu Ile Glu Leu Leu Ser Gln Asn Asp
1           5           10           15

```

```

Val Val Lys Tyr Gly Asp Ile Val Phe Arg Pro Leu Phe Lys Ser Ser
          20           25           30

```

```

Pro Glu Ala Gly Leu Leu Glu Leu Glu Lys Asn Gly Leu Ile Thr Ile
          35           40           45

```

```

Ser Arg Asn Arg Gly Val Leu Gln Asp Ile Arg Pro Ala Lys Pro Leu
          50           55           60

```

```

Phe Lys Ala Ala Phe Ser Tyr Leu Leu Gln Asp Lys Asp Leu Ser Ile
65           70           75           80

```

```

Val Leu Arg Thr Gly Tyr Tyr Leu Arg Leu Ile Ala Phe Glu Thr Gly
          85           90           95

```

```

Arg Ile Lys Lys Trp Glu Glu Glu Leu Arg Leu Leu Ala Lys Val Ser
          100          105          110

```

Asp Gln Arg Ile Cys Lys Ser Arg Leu Asn Tyr Leu Ala Ser Lys Ile
 115 120 125

Asp Ala Ser Ser Gly Val Ile Asn Ser Cys Glu Asp Lys Val Lys Glu
 130 135 140

Met Ser Lys
 145

<210> 12

<211> 4077

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (502) .. (2919)

<223>

<220>

<221> misc_feature

<223> Oligo 109

<400> 12

cctgtaccgg gtgcggtgcc agatacatcg caatcccctt cgtgctgttg ccatgcggga	60
cacaagcagg tccattccc atctgccac ctgcaccgc gccgagtaat gcttctcccc	120
cccgtagttc gctccaccgc tcgatccgcc acgtgaaccg gccctcgtac tcggtctctt	180
ctccttctag cggcggcaaa acctcctcga acgtcttgcc gagctccacc gcctccatgt	240
ttgtttcccg ggtttctggc acatccatta ctatcttgat atagttagat gttgatactg	300
cctgggtcca gcctactatc caagtcctgg tttgttcac gtgtgtagcc ctgtagagag	360
tgaatttcaa ggggcaggta acattgccgg acgcctacac ccgtggcttc tgaagaccat	420
cgcaaaaggt attcagaact ttgcaactag taccagcgat acaaggccgg tgacagcaga	480
agtaacgtcc cgagaaggcc a atg atg ttt ctg gag atg cag cga gcg ttc	531
Met Met Phe Leu Glu Met Gln Arg Ala Phe	
1 5 10	
atg cta cac ggg cgg cga gca gtg acg aga agt gcg gtc ggc gtg cgg	579

Met	Leu	His	Gly	Arg	Arg	Ala	Val	Thr	Arg	Ser	Ala	Val	Gly	Val	Arg	
				15					20					25		
tac	atc	tcg	gag	gac	atc	cag	cag	aag	gac	gca	cag	gcg	ggc	gag	aag	627
Tyr	Ile	Ser	Glu	Asp	Ile	Gln	Gln	Lys	Asp	Ala	Gln	Ala	Gly	Glu	Lys	
			30					35					40			
gcg	acg	gcg	acg	gcg	acg	ggg	gtt	atc	tac	aag	tcg	gac	gag	gaa	acg	675
Ala	Thr	Ala	Thr	Ala	Thr	Gly	Val	Ile	Tyr	Lys	Ser	Asp	Glu	Glu	Thr	
			45				50					55				
ctg	atg	tat	ttc	gac	aac	gta	tac	ccg	cgg	gcg	acg	tcg	ttg	tgg	cgg	723
Leu	Met	Tyr	Phe	Asp	Asn	Val	Tyr	Pro	Arg	Ala	Thr	Ser	Leu	Trp	Arg	
	60					65					70					
ccg	acg	cag	tgg	tac	aac	att	ctg	ctg	tcg	aac	cag	tcg	cgg	gaa	gcg	771
Pro	Thr	Gln	Trp	Tyr	Asn	Ile	Leu	Leu	Ser	Asn	Gln	Ser	Arg	Glu	Ala	
75					80					85					90	
gtg	agg	gag	aag	atc	atg	cgg	ttc	gca	agc	ccg	gca	agc	aac	ccg	gtg	819
Val	Arg	Glu	Lys	Ile	Met	Arg	Phe	Ala	Ser	Pro	Ala	Ser	Asn	Pro	Val	
				95					100					105		
cac	gga	ctg	gag	ctg	cgg	tcg	acg	atc	ccg	atc	aag	cgg	gac	ggc	ggg	867
His	Gly	Leu	Glu	Leu	Arg	Ser	Thr	Ile	Pro	Ile	Lys	Arg	Asp	Gly	Gly	
			110					115					120			
gtg	ttt	gcg	acg	ttc	cgg	gtg	ccg	cgt	gag	tac	acg	agg	gcg	cag	gtg	915
Val	Phe	Ala	Thr	Phe	Arg	Val	Pro	Arg	Glu	Tyr	Thr	Arg	Ala	Gln	Val	
			125				130					135				
aac	gcg	ttg	atc	cag	gcg	acc	acg	cag	cag	gag	tcg	tcg	aag	tcg	ctg	963
Asn	Ala	Leu	Ile	Gln	Ala	Thr	Thr	Gln	Gln	Glu	Ser	Ser	Lys	Ser	Leu	
		140				145					150					
ctg	gcg	gcg	ttt	acg	cgg	gct	gca	gcg	ttc	ccg	gtg	aag	ggc	gtg	ccg	1011
Leu	Ala	Ala	Phe	Thr	Arg	Ala	Ala	Ala	Phe	Pro	Val	Lys	Gly	Val	Pro	
155					160					165					170	
tgg	atc	gag	gac	ctg	aag	cgg	ctg	ccg	aac	aac	gtg	gtg	cgc	gtg	gag	1059
Trp	Ile	Glu	Asp	Leu	Lys	Arg	Leu	Pro	Asn	Asn	Val	Val	Arg	Val	Glu	
				175					180					185		
gtg	gag	ggc	ccg	gcg	ctg	tcg	gag	gag	gag	ttg	tac	tcg	ctg	ttc	cgg	1107
Val	Glu	Gly	Pro	Ala	Leu	Ser	Glu	Glu	Glu	Leu	Tyr	Ser	Leu	Phe	Arg	
			190					195					200			
cgg	tac	ggc	acg	att	ctg	gac	atc	tac	ccc	gcg	ggc	aag	aac	ggc	tat	1155
Arg	Tyr	Gly	Thr	Ile	Leu	Asp	Ile	Tyr	Pro	Ala	Gly	Lys	Asn	Gly	Tyr	
		205					210					215				
gcg	acg	atc	cgc	tac	cgc	tcg	ttc	cgg	ggc	gcg	atc	tgc	gcg	aag	aac	1203
Ala	Thr	Ile	Arg	Tyr	Arg	Ser	Phe	Arg	Gly	Ala	Ile	Cys	Ala	Lys	Asn	
		220				225					230					
tgt	gtg	tcg	ggg	atc	gag	atc	aac	ggc	tcg	acg	ctg	cat	gtc	aag	ttt	1251
Cys	Val	Ser	Gly	Ile	Glu	Ile	Asn	Gly	Ser	Thr	Leu	His	Val	Lys	Phe	
235					240				245						250	
gag	cct	gta	gtc	cgc	gcg	cac	gcg	atc	cgc	gac	ttc	ttc	gtg	aac	cac	1299
Glu	Pro	Val	Val	Arg	Ala	His	Ala	Ile	Arg	Asp	Phe	Phe	Val	Asn	His	
				255				260						265		

ccg cgc att gcc att ccc ctg ctg att gcc ctg tta tcc atc tgt gct	1347
Pro Arg Ile Ala Ile Pro Leu Leu Ile Ala Leu Leu Ser Ile Cys Ala	
270 275 280	
gtg ctg atc ttc gac cct atc cgg gag ttt tcc att gag cag aag atc	1395
Val Leu Ile Phe Asp Pro Ile Arg Glu Phe Ser Ile Glu Gln Lys Ile	
285 290 295	
acg cgc atg tac aca ctg tcg cgg gac aac ttc gtg gtg aaa agc atc	1443
Thr Arg Met Tyr Thr Leu Ser Arg Asp Asn Phe Val Val Lys Ser Ile	
300 305 310	
ctg cgc ctc acc tcc tat acg gtc tca tct gtg aag cat ctc tgg ggc	1491
Leu Arg Leu Thr Ser Tyr Thr Val Ser Ser Val Lys His Leu Trp Gly	
315 320 325 330	
tac gac gat gac cag cca gag aaa cga cag ttg tgg cag gag cgc gtg	1539
Tyr Asp Asp Asp Gln Pro Glu Lys Arg Gln Leu Trp Gln Glu Arg Val	
335 340 345	
gag aag gtg aac gac ctg aag atg tgg tta gag gaa aac aac aac act	1587
Glu Lys Val Asn Asp Leu Lys Met Trp Leu Glu Glu Asn Asn Asn Thr	
350 355 360	
ttc gtc gtg gtg acc ggt cct cgc ggc tct ggc aag cat gag ctc gtg	1635
Phe Val Val Val Thr Gly Pro Arg Gly Ser Gly Lys His Glu Leu Val	
365 370 375	
atg cag cac act ctg cat gac cgt cct aat gtg cta tat ctg gat tgt	1683
Met Gln His Thr Leu His Asp Arg Pro Asn Val Leu Tyr Leu Asp Cys	
380 385 390	
gat act ttg atc aag tcg cgt act gat tcc aag ttt ttg agg aat gcc	1731
Asp Thr Leu Ile Lys Ser Arg Thr Asp Ser Lys Phe Leu Arg Asn Ala	
395 400 405 410	
gcc cac cag att ggc tat ttc cca ata ttc ccg tgg ttg aat tcc gtc	1779
Ala His Gln Ile Gly Tyr Phe Pro Ile Phe Pro Trp Leu Asn Ser Val	
415 420 425	
acg act ctg gtt gac ctg gct gtt cag ggc ctg acc ggc cag aag tcc	1827
Thr Thr Leu Val Asp Leu Ala Val Gln Gly Leu Thr Gly Gln Lys Ser	
430 435 440	
ggg cta tct gag tca aag gaa acg caa ttc cgg aat atg ctg aac act	1875
Gly Leu Ser Glu Ser Lys Glu Thr Gln Phe Arg Asn Met Leu Asn Thr	
445 450 455	
gct atg atg tct att agg cac att gct ttg tcc ggt tat aag gcg act	1923
Ala Met Met Ser Ile Arg His Ile Ala Leu Ser Gly Tyr Lys Ala Thr	
460 465 470	
ctg cac tct ggc gat gat gtt acg act gtg aag gag gag gac tac ttg	1971
Leu His Ser Gly Asp Asp Val Thr Thr Val Lys Glu Glu Asp Tyr Leu	
475 480 485 490	
caa cag cat cct gag cgc aag ccg gtt att gta ata gac agg ttc agt	2019
Gln Gln His Pro Glu Arg Lys Pro Val Ile Val Ile Asp Arg Phe Ser	
495 500 505	
aac aag gct gag ata aac ggt ttt gtg tac aaa gaa ctc gca gat tgg	2067

Asn	Lys	Ala	Glu	Ile	Asn	Gly	Phe	Val	Tyr	Lys	Glu	Leu	Ala	Asp	Trp	
			510					515					520			
gcg	tcc	atg	ctt	gtt	cag	atg	aac	atc	gcg	cat	gtt	ata	ttc	ctc	acc	2115
Ala	Ser	Met	Leu	Val	Gln	Met	Asn	Ile	Ala	His	Val	Ile	Phe	Leu	Thr	
		525					530					535				
gag	tct	gta	tct	cca	aac	cag	ttg	ttg	gca	gag	gct	ctt	cca	aat	caa	2163
Glu	Ser	Val	Ser	Pro	Asn	Gln	Leu	Leu	Ala	Glu	Ala	Leu	Pro	Asn	Gln	
		540				545					550					
gtt	ttc	aag	ttc	ctg	ttt	ctc	tca	gat	gcg	tcg	aag	gac	tca	gcg	agg	2211
Val	Phe	Lys	Phe	Leu	Phe	Leu	Ser	Asp	Ala	Ser	Lys	Asp	Ser	Ala	Arg	
					560					565					570	
agc	tac	gtg	ctt	tca	cag	tta	tac	cct	tct	agc	cca	gcg	tat	tct	gaa	2259
Ser	Tyr	Val	Leu	Ser	Gln	Leu	Tyr	Pro	Ser	Ser	Pro	Ala	Tyr	Ser	Glu	
				575					580					585		
aaa	atg	cca	gca	gca	gat	gca	gat	gcc	aat	gag	gag	tat	agg	aag	gag	2307
Lys	Met	Pro	Ala	Ala	Asp	Ala	Asp	Ala	Asn	Glu	Glu	Tyr	Arg	Lys	Glu	
			590					595					600			
ata	gac	cgt	gca	ttg	gag	cct	atc	ggc	ggg	aga	atg	ttg	gac	cta	caa	2355
Ile	Asp	Arg	Ala	Leu	Glu	Pro	Ile	Gly	Gly	Arg	Met	Leu	Asp	Leu	Gln	
		605					610					615				
gcg	ttt	gtc	cgc	aga	gtg	aaa	tca	ggg	gag	gag	ccg	tct	gaa	gca	ttg	2403
Ala	Phe	Val	Arg	Arg	Val	Lys	Ser	Gly	Glu	Glu	Pro	Ser	Glu	Ala	Leu	
		620				625					630					
gaa	aag	atg	gtg	gag	caa	gca	tca	gag	cag	atc	act	caa	ata	ttc	tta	2451
Glu	Lys	Met	Val	Glu	Gln	Ala	Ser	Glu	Gln	Ile	Thr	Gln	Ile	Phe	Leu	
		635			640					645					650	
agt	gag	aga	tcg	gag	ccg	att	aaa	acg	gca	caa	gct	tgg	gag	ctg	atc	2499
Ser	Glu	Arg	Ser	Glu	Pro	Ile	Lys	Thr	Ala	Gln	Ala	Trp	Glu	Leu	Ile	
				655					660					665		
gaa	cta	ctt	tca	caa	aat	gat	gtc	gtg	aag	tat	gga	gat	atc	gtt	ttc	2547
Glu	Leu	Leu	Ser	Gln	Asn	Asp	Val	Val	Lys	Tyr	Gly	Asp	Ile	Val	Phe	
			670					675					680			
agg	ccg	ttg	ttt	aag	tct	tct	cca	gag	gcc	ggg	ttg	ttg	gag	tta	gag	2595
Arg	Pro	Leu	Phe	Lys	Ser	Ser	Pro	Glu	Ala	Gly	Leu	Leu	Glu	Leu	Glu	
		685					690					695				
aaa	aat	ggg	ctg	atc	aca	atc	agc	cgc	aat	aga	gga	gtg	ttg	caa	gat	2643
Lys	Asn	Gly	Leu	Ile	Thr	Ile	Ser	Arg	Asn	Arg	Gly	Val	Leu	Gln	Asp	
		700				705					710					
atc	cgg	ccc	gct	aaa	cca	ttg	ttt	aaa	gcc	gca	ttc	agc	tac	ctg	tta	2691
Ile	Arg	Pro	Ala	Lys	Pro	Leu	Phe	Lys	Ala	Ala	Phe	Ser	Tyr	Leu	Leu	
		715			720				725						730	
cag	gac	aag	gac	ctg	tcc	atc	gtc	ctc	cga	aca	ggg	tac	tac	ttg	cga	2739
Gln	Asp	Lys	Asp	Leu	Ser	Ile	Val	Leu	Arg	Thr	Gly	Tyr	Tyr	Leu	Arg	
				735					740					745		
cta	atc	gcg	ttt	gaa	act	ggg	cgg	atc	aaa	aag	tgg	gag	gag	gaa	ttg	2787
Leu	Ile	Ala	Phe	Glu	Thr	Gly	Arg	Ile	Lys	Lys	Trp	Glu	Glu	Glu	Leu	
			750					755						760		

```

cgt cta ttg gca aag gtt agt gac caa aga ata tgc aag agt cgg ttg      2835
Arg Leu Leu Ala Lys Val Ser Asp Gln Arg Ile Cys Lys Ser Arg Leu
      765                      770                      775

aat tac tta gca agc aaa ata gac gcc agt agt ggg gta att aac agt      2883
Asn Tyr Leu Ala Ser Lys Ile Asp Ala Ser Ser Gly Val Ile Asn Ser
      780                      785                      790

tgt gag gac aag gta aaa gaa atg tca aag cgt att tagaaacgct          2929
Cys Glu Asp Lys Val Lys Glu Met Ser Lys Arg Ile
      795                      800                      805

cgcatcaata agacccatt ttgccacgat ctggcaagat gaactaaata gcgataatca      2989

cactaggtct cattagttta gcattccagc tgcacgggct ttagtttagt acctcaattc      3049

caaaaacggc actatcttgt ccgagcgctg gtgaatatcc ttgccgtatg tgaaattatc      3109

cagggttgat gtagttggcc ggagctacca gtaattgttg gcctatgcac gggccgctat      3169

tttccgtaac ataatccaga gatccgttaa caggcaacat gaactgattg attaatttta      3229

atgaactttt ttttttggct ttgaaaggcc taaaattgga tcatctacat tccctgacga      3289

tataggatgg taaaagccct accaaccaag aaattctgat ccctttgaca aatcccttat      3349

aacatagata ttagaatata atcaacccca ttgtgtgaca actcttctcc aaaaataatt      3409

ggtctttcac acaattaaag attggcttca aggtagctcc acggtgagat ctgctaaaac      3469

agtatggaag ctgctaccat taatgaagag caagtcattc ctgaaaggga tttggaaaag      3529

ctgccagtgt cgagaaatgt cagtagagga tgggcagttc gtaagttggc aatattatac      3589

tttctattag cgtctttcgc tatagggacg gtgcacttat tccagtctgg ttacatgggt      3649

gctaatagta aatctcctca gatcggtcgc tggcaactat ttgcggagaa ttctcagggc      3709

cttacaaaac ggagtagatg ggcgatgatc gctgaaaaac accacgcgcc cctcttagca      3769

gcggcaagtc tatcaacggc agtggagcct aatgtttgcc cgctgatgg cagaagatgc      3829

atgggcgaac taatcaccag attattactt caatcactcg cgatagctca ggacctctat      3889

ggtggtgagt ggaagattac tcagccacat ccaccatata acaacaacaa acgcagtctg      3949

tcgaactatg catttgaaac tgcatcggaa ctttcaactc aaataaagag tagagcagac      4009

actttggcgg cgtaaataat caatctggtg ggacgcttcc gcagcaatga cccaaatggc      4069

cttgcgaa                                          4077

```

<210> 13

<211> 806

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 109

<400> 13

Met Met Phe Leu Glu Met Gln Arg Ala Phe Met Leu His Gly Arg Arg
 1 5 10 15

Ala Val Thr Arg Ser Ala Val Gly Val Arg Tyr Ile Ser Glu Asp Ile
 20 25 30

Gln Gln Lys Asp Ala Gln Ala Gly Glu Lys Ala Thr Ala Thr Ala Thr
 35 40 45

Gly Val Ile Tyr Lys Ser Asp Glu Glu Thr Leu Met Tyr Phe Asp Asn
 50 55 60

Val Tyr Pro Arg Ala Thr Ser Leu Trp Arg Pro Thr Gln Trp Tyr Asn
 65 70 75 80

Ile Leu Leu Ser Asn Gln Ser Arg Glu Ala Val Arg Glu Lys Ile Met
 85 90 95

Arg Phe Ala Ser Pro Ala Ser Asn Pro Val His Gly Leu Glu Leu Arg
 100 105 110

Ser Thr Ile Pro Ile Lys Arg Asp Gly Gly Val Phe Ala Thr Phe Arg
 115 120 125

Val Pro Arg Glu Tyr Thr Arg Ala Gln Val Asn Ala Leu Ile Gln Ala
 130 135 140

Thr Thr Gln Gln Glu Ser Ser Lys Ser Leu Leu Ala Ala Phe Thr Arg
 145 150 155 160

Ala Ala Ala Phe Pro Val Lys Gly Val Pro Trp Ile Glu Asp Leu Lys
 165 170 175

Arg Leu Pro Asn Asn Val Val Arg Val Glu Val Glu Gly Pro Ala Leu
 180 185 190

Ser Glu Glu Glu Leu Tyr Ser Leu Phe Arg Arg Tyr Gly Thr Ile Leu
 195 200 205

Asp Ile Tyr Pro Ala Gly Lys Asn Gly Tyr Ala Thr Ile Arg Tyr Arg
 210 215 220

Ser Phe Arg Gly Ala Ile Cys Ala Lys Asn Cys Val Ser Gly Ile Glu
 225 230 235 240

Ile Asn Gly Ser Thr Leu His Val Lys Phe Glu Pro Val Val Arg Ala
 245 250 255

His Ala Ile Arg Asp Phe Phe Val Asn His Pro Arg Ile Ala Ile Pro
 260 265 270

Leu Leu Ile Ala Leu Leu Ser Ile Cys Ala Val Leu Ile Phe Asp Pro
 275 280 285

Ile Arg Glu Phe Ser Ile Glu Gln Lys Ile Thr Arg Met Tyr Thr Leu
 290 295 300

Ser Arg Asp Asn Phe Val Val Lys Ser Ile Leu Arg Leu Thr Ser Tyr
 305 310 315 320

Thr Val Ser Ser Val Lys His Leu Trp Gly Tyr Asp Asp Asp Gln Pro
 325 330 335

Glu Lys Arg Gln Leu Trp Gln Glu Arg Val Glu Lys Val Asn Asp Leu
 340 345 350

Lys Met Trp Leu Glu Glu Asn Asn Asn Thr Phe Val Val Val Thr Gly
 355 360 365

Pro Arg Gly Ser Gly Lys His Glu Leu Val Met Gln His Thr Leu His
 370 375 380

Asp Arg Pro Asn Val Leu Tyr Leu Asp Cys Asp Thr Leu Ile Lys Ser
 385 390 395 400

Arg Thr Asp Ser Lys Phe Leu Arg Asn Ala Ala His Gln Ile Gly Tyr
 405 410 415

Phe Pro Ile Phe Pro Trp Leu Asn Ser Val Thr Thr Leu Val Asp Leu
 420 425 430

Ala Val Gln Gly Leu Thr Gly Gln Lys Ser Gly Leu Ser Glu Ser Lys
 435 440 445

Glu Thr Gln Phe Arg Asn Met Leu Asn Thr Ala Met Met Ser Ile Arg

450		455		460
His Ile Ala Leu Ser Gly Tyr Lys Ala Thr Leu His Ser Gly Asp Asp				
465		470		475 480
Val Thr Thr Val Lys Glu Glu Asp Tyr Leu Gln Gln His Pro Glu Arg				
	485		490	495
Lys Pro Val Ile Val Ile Asp Arg Phe Ser Asn Lys Ala Glu Ile Asn				
	500		505	510
Gly Phe Val Tyr Lys Glu Leu Ala Asp Trp Ala Ser Met Leu Val Gln				
	515		520	525
Met Asn Ile Ala His Val Ile Phe Leu Thr Glu Ser Val Ser Pro Asn				
	530		535	540
Gln Leu Leu Ala Glu Ala Leu Pro Asn Gln Val Phe Lys Phe Leu Phe				
545		550		555 560
Leu Ser Asp Ala Ser Lys Asp Ser Ala Arg Ser Tyr Val Leu Ser Gln				
	565		570	575
Leu Tyr Pro Ser Ser Pro Ala Tyr Ser Glu Lys Met Pro Ala Ala Asp				
	580		585	590
Ala Asp Ala Asn Glu Glu Tyr Arg Lys Glu Ile Asp Arg Ala Leu Glu				
	595		600	605
Pro Ile Gly Gly Arg Met Leu Asp Leu Gln Ala Phe Val Arg Arg Val				
	610		615	620
Lys Ser Gly Glu Glu Pro Ser Glu Ala Leu Glu Lys Met Val Glu Gln				
625		630		635 640
Ala Ser Glu Gln Ile Thr Gln Ile Phe Leu Ser Glu Arg Ser Glu Pro				
	645		650	655
Ile Lys Thr Ala Gln Ala Trp Glu Leu Ile Glu Leu Leu Ser Gln Asn				
	660		665	670
Asp Val Val Lys Tyr Gly Asp Ile Val Phe Arg Pro Leu Phe Lys Ser				
	675		680	685
Ser Pro Glu Ala Gly Leu Leu Glu Leu Glu Lys Asn Gly Leu Ile Thr				
	690		695	700

Ile Ser Arg Asn Arg Gly Val Leu Gln Asp Ile Arg Pro Ala Lys Pro
705 710 715 720

Leu Phe Lys Ala Ala Phe Ser Tyr Leu Leu Gln Asp Lys Asp Leu Ser
725 730 735

Ile Val Leu Arg Thr Gly Tyr Tyr Leu Arg Leu Ile Ala Phe Glu Thr
740 745 750

Gly Arg Ile Lys Lys Trp Glu Glu Glu Leu Arg Leu Leu Ala Lys Val
755 760 765

Ser Asp Gln Arg Ile Cys Lys Ser Arg Leu Asn Tyr Leu Ala Ser Lys
770 775 780

Ile Asp Ala Ser Ser Gly Val Ile Asn Ser Cys Glu Asp Lys Val Lys
785 790 795 800

Glu Met Ser Lys Arg Ile
805

<210> 14

<211> 1337

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 163

<400> 14

gatcttggac ttgacacaga cagtggcacg ttacggctgg ttaccattca tcctttacat	60
gggctggctg cacttgcca actccccaa cctgctaaac cttctctccc cactcccaag	120
tgtctaattgt atttgcaacg gttccctcag tatgtgcttt tgttcacatg atcctcctag	180
ggcaaggccg ggacggaggg aacgatagcg cgaggctacc agcctcggtc ctaccacgtc	240
gagtgtgccg caattctagt gagcttgat atacgtaatt tgaataaata agcgttcatt	300
gttaccttat ccgcctatga gttcgggaac tacctgttcg ctgcctttgt gaaaagtagc	360
tgacaatagc agcggtgatg catcttccga aacagccata ggagagcaat gggccacgag	420

```

ctgacaagcc tgcagcgtga agtgatttcc ggactcacag caggaacaat cacaacaatc 480
gctagccacc cgctcgacct gcttaaactg cgactacaat tatcagcggg caacagggct 540
aacactacat atacaggact aatcagagac atatttgaac gccagcaatg gggacgagag 600
ctatatcgtg gtttaggcgt taacctgctt ggaaattcag tggcatgggc gctttacttt 660
ggctgttacc gatgcgcaaa ggacattgcg cttcgacact tgggcaatga gtctgctacg 720
ggatatcatgg accgcgcct gccagcgcac gcatacatgc ttgcagctgg tagcagcggg 780
attgcgacgg cagtccttac aaaccccata tgggtcatta agactcgcat aatggccact 840
tctcgtgctg gaccttaciaa gtcgacgttt gacgggggttt ataagttata tcaaactgaa 900
gggtgttctg cattctggcg ggggtgtgtt ccatcgctac tcggcgtctc gcaaggagct 960
atctactttg cgctgtatga tacattgaaa ttccattacc tgcactctag tactgacaag 1020
gccgagcgaa ggttgtcggg ttccgagatc atcggcataa catgtatctc taaaatgatt 1080
tccgtcacat cggcttatcc atttcagctg ctgaaatcta agctgcagga ttttgggtgca 1140
ccatccggca tcacccaact tgttcagact gtttacagta gagagggtat cagaggcttc 1200
tacagggggc tatcctgcta atctcctgcg ggcagttcct gctacatgca taaccttttt 1260
cgtgtacgaa aatatcaaat atcgtctcta aatttgtgat ttggtgcagc gcgtagcaag 1320
tatacactga ctggatc 1337

```

<210> 15

<211> 267

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 163

<400> 15

```

His Glu Leu Thr Ser Leu Gln Arg Glu Val Ile Ser Gly Leu Thr Ala
1           5           10           15

```

```

Gly Thr Ile Thr Thr Ile Ala Ser His Pro Leu Asp Leu Leu Lys Leu
20           25           30

```

Arg Leu Gln Leu Ser Ala Gly Asn Arg Ala Asn Thr Thr Tyr Thr Gly
 35 40 45
 Leu Ile Arg Asp Ile Phe Glu Arg Gln Gln Trp Gly Arg Glu Leu Tyr
 50 55 60
 Arg Gly Leu Gly Val Asn Leu Leu Gly Asn Ser Val Ala Trp Ala Leu
 65 70 75 80
 Tyr Phe Gly Cys Tyr Arg Cys Ala Lys Asp Ile Ala Leu Arg His Leu
 85 90 95
 Gly Asn Glu Ser Ala Thr Gly Ile Met Asp Arg Arg Leu Pro Ala His
 100 105 110
 Ala Tyr Met Leu Ala Ala Gly Ser Ser Gly Ile Ala Thr Ala Val Leu
 115 120 125
 Thr Asn Pro Ile Trp Val Ile Lys Thr Arg Ile Met Ala Thr Ser Arg
 130 135 140
 Ala Gly Pro Tyr Lys Ser Thr Phe Asp Gly Val Tyr Lys Leu Tyr Gln
 145 150 155 160
 Thr Glu Gly Val Leu Ala Phe Trp Arg Gly Val Val Pro Ser Leu Leu
 165 170 175
 Gly Val Ser Gln Gly Ala Ile Tyr Phe Ala Leu Tyr Asp Thr Leu Lys
 180 185 190
 Phe His Tyr Leu His Ser Ser Thr Asp Lys Ala Glu Arg Arg Leu Ser
 195 200 205
 Val Ser Glu Ile Ile Gly Ile Thr Cys Ile Ser Lys Met Ile Ser Val
 210 215 220
 Thr Ser Val Tyr Pro Phe Gln Leu Leu Lys Ser Lys Leu Gln Asp Phe
 225 230 235 240
 Gly Ala Pro Ser Gly Ile Thr Gln Leu Val Gln Thr Val Tyr Ser Arg
 245 250 255
 Glu Gly Ile Arg Gly Phe Tyr Arg Gly Leu Ser
 260 265

<211> 47

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 163

<400> 16

Ala	Ser	Pro	Asn	Leu	Phe	Arg	Leu	Phe	Thr	Val	Glu	Arg	Val	Ser	Glu
1				5					10					15	

Ala	Ser	Thr	Gly	Ala	Tyr	Pro	Ala	Asn	Leu	Leu	Arg	Ala	Val	Pro	Ala
			20					25					30		

Thr	Cys	Ile	Thr	Phe	Phe	Val	Tyr	Glu	Asn	Ile	Lys	Tyr	Arg	Leu
		35					40					45		

<210> 17

<211> 1733

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (329)..(1207)

<223>

<220>

<221> misc_feature

<223> Oligo 163

<400> 17	
actcccccaa cctgataaac cttctctccc cactcccaag tgtctaattgt atttgcaacg	60
gttccctcag tatgtgcttt tggtcacatg atcctcctag ggcaaggccg ggacggaggg	120

aacgatagcg cgaggctacc agcctcggtc ctaccacgtc gagtgtgccg caattctagt	180
gagcttgtat atacagtaat ttgaataaat aagcgttcat tgttacctta tccgcctatg	240
agttcgggaa ctactgttcg ctgcctttgt gaaaagtagc tgacaatagc agcgggatg	300
catcttccga aacagccata ggagagca atg ggc cac gag ctg aca agc ctg	352
Met Gly His Glu Leu Thr Ser Leu	
1 5	
cag cgt gaa gtg att tcc gga ctc aca gca gga aca atc aca aca atc	400
Gln Arg Glu Val Ile Ser Gly Leu Thr Ala Gly Thr Ile Thr Thr Ile	
10 15 20	
gct agc cac ccg ctc gac ctg ctt aaa ctg cga cta caa tta tca gcg	448
Ala Ser His Pro Leu Asp Leu Leu Lys Leu Arg Leu Gln Leu Ser Ala	
25 30 35 40	
ggc aac agg gct aac act aca tat aca gga cta atc aga gac ata ttt	496
Gly Asn Arg Ala Asn Thr Thr Tyr Thr Gly Leu Ile Arg Asp Ile Phe	
45 50 55	
gaa cgc cag caa tgg gga cga gag cta tat cgt ggt tta ggc gtt aac	544
Glu Arg Gln Gln Trp Gly Arg Glu Leu Tyr Arg Gly Leu Gly Val Asn	
60 65 70	
ctg ctt gga aat tca gtg gca tgg gcg ctt tac ttt ggc tgt tac cga	592
Leu Leu Gly Asn Ser Val Ala Trp Ala Leu Tyr Phe Gly Cys Tyr Arg	
75 80 85	
tgc gca aag gac att gcg ctt cga cac ttg ggc aat gag tct gct acg	640
Cys Ala Lys Asp Ile Ala Leu Arg His Leu Gly Asn Glu Ser Ala Thr	
90 95 100	
ggt atc atg gac cgc cgc ctg cca gcg cac gca tac atg ctt gca gct	688
Gly Ile Met Asp Arg Arg Leu Pro Ala His Ala Tyr Met Leu Ala Ala	
105 110 115 120	
ggt agc agc ggg att gcg acg gca gtc ctt aca aac ccc ata tgg gtc	736
Gly Ser Ser Gly Ile Ala Thr Ala Val Leu Thr Asn Pro Ile Trp Val	
125 130 135	
att aag act cgc ata atg gcc act tct cgt gct gga cct tac aag tcg	784
Ile Lys Thr Arg Ile Met Ala Thr Ser Arg Ala Gly Pro Tyr Lys Ser	
140 145 150	
acg ttt gac ggg gtt tat aag tta tat caa act gaa ggt gtt ctg gca	832
Thr Phe Asp Gly Val Tyr Lys Leu Tyr Gln Thr Glu Gly Val Leu Ala	
155 160 165	
ttc tgg cgg ggt gtt gtt cca tcg cta ctc ggc gtc tcg caa gga gct	880
Phe Trp Arg Gly Val Val Pro Ser Leu Leu Gly Val Ser Gln Gly Ala	
170 175 180	
atc tac ttt gcg ctg tat gat aca ttg aaa ttc cat tac ctg cac tct	928
Ile Tyr Phe Ala Leu Tyr Asp Thr Leu Lys Phe His Tyr Leu His Ser	
185 190 195 200	
agt act gac aag gcc gag cga agg ttg tcg gtt tcc gag atc atc ggc	976
Ser Thr Asp Lys Ala Glu Arg Arg Leu Ser Val Ser Glu Ile Ile Gly	
205 210 215	

ata aca tgt atc tct aaa atg att tcc gtc aca tcg gtc tat cca ttt 1024
 Ile Thr Cys Ile Ser Lys Met Ile Ser Val Thr Ser Val Tyr Pro Phe
 220 225 230

cag ctg ctg aaa tct aag ctg cag gat ttt ggt gca cca tcc ggc atc 1072
 Gln Leu Leu Lys Ser Lys Leu Gln Asp Phe Gly Ala Pro Ser Gly Ile
 235 240 245

acc caa ctt gtt cag act gtt tac agt aga gag ggt atc aga ggc ttc 1120
 Thr Gln Leu Val Gln Thr Val Tyr Ser Arg Glu Gly Ile Arg Gly Phe
 250 255 260

tac agg ggc cta tcc gct aat ctc ctg cgg gca gtt cct gct aca tgc 1168
 Tyr Arg Gly Leu Ser Ala Asn Leu Leu Arg Ala Val Pro Ala Thr Cys
 265 270 275 280

ata acc ttt ttc gtg tac gaa aat atc aaa tat cgt ctc taaatttggtg 1217
 Ile Thr Phe Phe Val Tyr Glu Asn Ile Lys Tyr Arg Leu
 285 290

atttggtgca ggcgcgtagca agtatacaact gactggatct aaaagatgat taatttgtct 1277

taagcttttac aaaaaaacac tttatatatt gattgatatt gggcaaaaga ttcattgacct 1337

tatgatgaaa tttagtaacc gagcttggcc aaatcctcag cgatctcgga cttagcatca 1397

tatttagcct tgacagcagc agccggttttg ttgaaggccc tcttcttgga gtagtactca 1457

gcagatctga cctttctctt ctcttccaac ttggcaacaa cgttctcgta cttccagcca 1517

acggtggaag acaacttacc caaggtggtg aactttctac ctggcttcaa tctcaaaact 1577

ctaaagcctg tggaacaaca actctcttca acttgctgta tggaggtggg acaccctcta 1637

agacctttaa tctctccata gcagccttac cacggggcag tcttgtgagc aaccatacct 1697

ctgacagcct tgtagaagaa tctggatggg gctctg 1733

<210> 18

<211> 293

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 163

<400> 18

Met Gly His Glu Leu Thr Ser Leu Gln Arg Glu Val Ile Ser Gly Leu
 1 5 10 15

Thr Ala Gly Thr Ile Thr Thr Ile Ala Ser His Pro Leu Asp Leu Leu
 20 25 30
 Lys Leu Arg Leu Gln Leu Ser Ala Gly Asn Arg Ala Asn Thr Thr Tyr
 35 40 45
 Thr Gly Leu Ile Arg Asp Ile Phe Glu Arg Gln Gln Trp Gly Arg Glu
 50 55 60
 Leu Tyr Arg Gly Leu Gly Val Asn Leu Leu Gly Asn Ser Val Ala Trp
 65 70 75 80
 Ala Leu Tyr Phe Gly Cys Tyr Arg Cys Ala Lys Asp Ile Ala Leu Arg
 85 90 95
 His Leu Gly Asn Glu Ser Ala Thr Gly Ile Met Asp Arg Arg Leu Pro
 100 105 110
 Ala His Ala Tyr Met Leu Ala Ala Gly Ser Ser Gly Ile Ala Thr Ala
 115 120 125
 Val Leu Thr Asn Pro Ile Trp Val Ile Lys Thr Arg Ile Met Ala Thr
 130 135 140
 Ser Arg Ala Gly Pro Tyr Lys Ser Thr Phe Asp Gly Val Tyr Lys Leu
 145 150 155 160
 Tyr Gln Thr Glu Gly Val Leu Ala Phe Trp Arg Gly Val Val Pro Ser
 165 170 175
 Leu Leu Gly Val Ser Gln Gly Ala Ile Tyr Phe Ala Leu Tyr Asp Thr
 180 185 190
 Leu Lys Phe His Tyr Leu His Ser Ser Thr Asp Lys Ala Glu Arg Arg
 195 200 205
 Leu Ser Val Ser Glu Ile Ile Gly Ile Thr Cys Ile Ser Lys Met Ile
 210 215 220
 Ser Val Thr Ser Val Tyr Pro Phe Gln Leu Leu Lys Ser Lys Leu Gln
 225 230 235 240
 Asp Phe Gly Ala Pro Ser Gly Ile Thr Gln Leu Val Gln Thr Val Tyr
 245 250 255
 Ser Arg Glu Gly Ile Arg Gly Phe Tyr Arg Gly Leu Ser Ala Asn Leu
 260 265 270

Leu Arg Ala Val Pro Ala Thr Cys Ile Thr Phe Phe Val Tyr Glu Asn
 275 280 285

Ile Lys Tyr Arg Leu
 290

<210> 19

<211> 972

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 31

<400> 19

gatcgatatt gtcggcaact tcattcctgg attaatatatt atgcaatcaa tatttgggta	60
cctatcgtgg gcaattatatt acaagtgggc aaaagattgg attaaagatg agctgcctgc	120
gcctgggttta ttaaacaatgt tgattaacat gtttttatct cccggagttg tcgatgaaaa	180
actatataca ggccagagct ttcttcaagt tctccttttg ctagctgctc tagtttgtgt	240
tccgtggcta ctactttaca aaccgttgat gttgaagcgc cagaacgata tagctctcag	300
caaaggattc agaagcctca gagaccaacg ggtgcatgag attcttcttg aagcacagga	360
aaacgcaggc gaagatatgc tggttgcaga ttatgaaaat gaagatgaat cgtcggagga	420
gttcaacttt ggtgacgtta tgatacatca gggtatccac accattgaat tttgtttgaa	480
ttgtatatct catactgcat cgtatctcag attgtgggccc ttatctcttg cgcattgctca	540
actctcgact gtggttggtt ctatgaccat tcagaattcg ttctccgact ccaaccctgg	600
ttcatttttc tctgtcacca aggtgggtgg tttggttgcc atgtgggttg tgttgactgt	660
ttgcatttta gtcttaatgg aaggaacgct ggctatgttg cactcggtta gattgcattg	720
ggtggaagcg atgtccaaat tcttcgaggg cgaaggctat gcttacgaac cattttcctt	780
caaggctatc aacagcgatg atgaatagta tcgtatatat taaaactaga cagatggagg	840
tagtgtcaca tgtgtcaggt atagattgcc cgatgaataa gcatccttga aatttagaaa	900
aatcataggc ctttagaatc gacatacaac gtctaaatat atttactatt cacttcatag	960
ttcatcgtga tc	972

<210> 20

<211> 263

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 31

<400> 20

Ile Asp Ile Val Gly Asn Phe Ile Pro Gly Leu Ile Phe Met Gln Ser
1 5 10 15

Ile Phe Gly Tyr Leu Ser Trp Ala Ile Ile Tyr Lys Trp Ser Lys Asp
20 25 30

Trp Ile Lys Asp Glu Leu Pro Ala Pro Gly Leu Leu Asn Met Leu Ile
35 40 45

Asn Met Phe Leu Ser Pro Gly Val Val Asp Glu Lys Leu Tyr Thr Gly
50 55 60

Gln Ser Phe Leu Gln Val Ile Leu Leu Leu Ala Ala Leu Val Cys Val
65 70 75 80

Pro Trp Leu Leu Leu Tyr Lys Pro Leu Met Leu Lys Arg Gln Asn Asp
85 90 95

Ile Ala Leu Ser Lys Gly Phe Arg Ser Leu Arg Asp Gln Arg Val His
100 105 110

Glu Ile Leu Leu Glu Ala Gln Glu Asn Ala Gly Glu Asp Met Leu Val
115 120 125

Ala Asp Tyr Glu Asn Glu Asp Glu Ser Ser Glu Glu Phe Asn Phe Gly
130 135 140

Asp Val Met Ile His Gln Val Ile His Thr Ile Glu Phe Cys Leu Asn
145 150 155 160

Cys Ile Ser His Thr Ala Ser Tyr Leu Arg Leu Trp Ala Leu Ser Leu
 165 170 175

Ala His Ala Gln Leu Ser Thr Val Leu Trp Ser Met Thr Ile Gln Asn
 180 185 190

Ser Phe Ser Asp Ser Asn Pro Gly Ser Phe Phe Ser Val Thr Lys Val
 195 200 205

Val Val Leu Phe Ala Met Trp Phe Val Leu Thr Val Cys Ile Leu Val
 210 215 220

Leu Met Glu Gly Thr Ser Ala Met Leu His Ser Leu Arg Leu His Trp
 225 230 235 240

Val Glu Ala Met Ser Lys Phe Phe Glu Gly Glu Gly Tyr Ala Tyr Glu
 245 250 255

Pro Phe Ser Phe Lys Ala Ile
 260

<210> 21

<211> 4054

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (623)..(3253)

<223>

<220>

<221> misc_feature

<223> Oligo 31

<400> 21
 ccgtgcagta catcttgccg attgccagta taggtcacc accactgtcg tcgcctgctt 60
 ctctgtatacc cactgttcca caacatacag ctctgtaccca ccaagtgcc cttcttggtt 120
 cgaacaatcc tccaaagcgc tctgcgacct gaaaaccgcc cggaagtccg gcacaacgta 180

ctttatccat cgcgcattgtg tttcggcctc accatcgctc accactttgc gtgaacctcg	240
agagttatgc tgggcaatat ccatatctaa gctccatcta ctactgttcg ctacctcacg	300
ggctgcctag gaagtggctc atagacgata tggagctata gggctcatcg cagtccgtca	360
taaggcatgt tcaccatcac agaaaacgat ttaaaaggga caccacaaac ttcaaccact	420
acaaatacca gtaaggatat cgctaggaag cagttggagt tactagctgt taggagacat	480
tgaacgagt atctaaagat tttgacagtg gggattctgt gtgcattggg gttcatgcac	540
gtagtgacta agtagtggct gcatacatag ggtacacaga ctgacagggt ggctgacaaa	600
agaacagtaa cttgagatcg tg atg ttc cta tca atg cgg gcc gag gag gcc	652
Met Phe Leu Ser Met Arg Ala Glu Glu Ala	
1 5 10	
atg ttt cga tct gcg gat atg acc tat atc gag ttg tac att ccg cta	700
Met Phe Arg Ser Ala Asp Met Thr Tyr Ile Glu Leu Tyr Ile Pro Leu	
15 20 25	
gag ata gcg cgc gag gtt gtt tgt gta tta ggt aac ctg ggg agc gta	748
Glu Ile Ala Arg Glu Val Val Cys Val Leu Gly Asn Leu Gly Ser Val	
30 35 40	
atg tta aag gac atg aac aaa gac cta agc acg ttc cag cgc ggg tac	796
Met Leu Lys Asp Met Asn Lys Asp Leu Ser Thr Phe Gln Arg Gly Tyr	
45 50 55	
gtt aac cag gtg cgg agg ttc gat gag gtc gag cgg cag gtg ggg tac	844
Val Asn Gln Val Arg Arg Phe Asp Glu Val Glu Arg Gln Val Gly Tyr	
60 65 70	
atg gag ggt gtg gta cgg agg cac aag aac gag aca tgg cgg tac ctg	892
Met Glu Gly Val Val Arg Arg His Lys Asn Glu Thr Trp Arg Tyr Leu	
75 80 85 90	
tat cga cat cta cag cgg gaa gag cag cag gag tat ccc gcc cgg gag	940
Tyr Arg His Leu Gln Arg Glu Glu Gln Gln Glu Tyr Pro Gly Arg Glu	
95 100 105	
cac ccc acg ctt gcg cag ctg atc ggt tca atg cac aca cac tcg att	988
His Pro Thr Leu Ala Gln Leu Ile Gly Ser Met His Thr His Ser Ile	
110 115 120	
gat tcg gtg gac gag gtt gcg gag gaa atc atg cag ttc gag ggg cgc	1036
Asp Ser Val Asp Glu Val Ala Glu Glu Ile Met Gln Phe Glu Gly Arg	
125 130 135	
gtg agg cag ctc gac cag agt ctt gta gcg atg cgc gag cga ctg tcg	1084
Val Arg Gln Leu Asp Gln Ser Leu Val Ala Met Arg Glu Arg Leu Ser	
140 145 150	
aag ctt gtg cat gag cga cgt gtg atg ttc act tgc gaa cac ttt ctc	1132
Lys Leu Val His Glu Arg Arg Val Met Phe Thr Cys Glu His Phe Leu	
155 160 165 170	
gaa gtg aac ccc ggg atc ggg gag cga ttg ccc aca cgc ccc gcg ggg	1180
Glu Val Asn Pro Gly Ile Gly Glu Arg Leu Pro Thr Arg Pro Ala Gly	
175 180 185	

ttc gaa gct gat gag ttt gaa ctc acg cgt gtc ggg gag gag gac gag	1228
Phe Glu Ala Asp Glu Phe Glu Leu Thr Arg Val Gly Glu Glu Asp Glu	
190 195 200	
gag acg gca agc cag ctg tct ttt gat att tcg gat gac gca gag aca	1276
Glu Thr Ala Ser Gln Leu Ser Phe Asp Ile Ser Asp Asp Ala Glu Thr	
205 210 215	
cag ctg ccc ggg gac atg cgc acg tta ctc gaa ccg gtg tac cgg cat	1324
Gln Leu Pro Gly Asp Met Arg Thr Leu Leu Glu Pro Val Tyr Arg His	
220 225 230	
cag tac cta ctt aca ggc tca att gag cgt gct aag gta gag gcg ctc	1372
Gln Tyr Leu Leu Thr Gly Ser Ile Glu Arg Ala Lys Val Glu Ala Leu	
235 240 245 250	
aac aag atc ttg tgg cgc ctt ctc cgt ggg aat gtt ttc ttc cag aat	1420
Asn Lys Ile Leu Trp Arg Leu Leu Arg Gly Asn Val Phe Phe Gln Asn	
255 260 265	
ttc cct gtc tca gtg tca ccg gtg gaa gaa gat gat acc gat ctc gag	1468
Phe Pro Val Ser Val Ser Pro Val Glu Glu Asp Asp Thr Asp Leu Glu	
270 275 280	
act gac tgc ttt atc gtc ttt acg cat ggc gaa gtc ttg cta agc aag	1516
Thr Asp Cys Phe Ile Val Phe Thr His Gly Glu Val Leu Leu Ser Lys	
285 290 295	
gcg aag aaa gtg ata gaa tcc cta aat ggc aca ata tat ccg ttt atg	1564
Ala Lys Lys Val Ile Glu Ser Leu Asn Gly Thr Ile Tyr Pro Phe Met	
300 305 310	
caa gac ggg gcg aca gtg cag gag ttg aac gac aag ata gcg gac ctt	1612
Gln Asp Gly Ala Thr Val Gln Glu Leu Asn Asp Lys Ile Ala Asp Leu	
315 320 325 330	
aag cag ata tgt tct acg aca gaa cag aca cta cat acg gaa ctt ttt	1660
Lys Gln Ile Cys Ser Thr Thr Glu Gln Thr Leu His Thr Glu Leu Phe	
335 340 345	
ctt gtt gcc aac caa ttg ccc atg tgg aat gcc att att aag cgt gaa	1708
Leu Val Ala Asn Gln Leu Pro Met Trp Asn Ala Ile Ile Lys Arg Glu	
350 355 360	
aag tac atc tac tct gcc ttg aat tta ttt agg cag gag tca cag ggc	1756
Lys Tyr Ile Tyr Ser Ala Leu Asn Leu Phe Arg Gln Glu Ser Gln Gly	
365 370 375	
ctc gtt gca gag gga tgg ctt ccc acg tac gat cta cca gga gtt cag	1804
Leu Val Ala Glu Gly Trp Leu Pro Thr Tyr Asp Leu Pro Gly Val Gln	
380 385 390	
gcg gca cta aag gac tat ggg gag agc gta gga tcc gca aat tcg gcc	1852
Ala Ala Leu Lys Asp Tyr Gly Glu Ser Val Gly Ser Ala Asn Ser Ala	
395 400 405 410	
gtt ttg aat gta att tcg aca acg agg aca ccg cca act ttc cac agg	1900
Val Leu Asn Val Ile Ser Thr Thr Arg Thr Pro Pro Thr Phe His Arg	
415 420 425	
act aac aag ttt acg cag gca ttc cag tcc att gtt gat gcc tat ggt	1948

Thr	Asn	Lys	Phe	Thr	Gln	Ala	Phe	Gln	Ser	Ile	Val	Asp	Ala	Tyr	Gly	
			430					435					440			
att	gca	aca	tat	aaa	gaa	gtg	aac	ccc	ggg	ttg	gca	act	atc	gtc	aca	1996
Ile	Ala	Thr	Tyr	Lys	Glu	Val	Asn	Pro	Gly	Leu	Ala	Thr	Ile	Val	Thr	
		445					450					455				
ttt	ccc	ttt	atg	ttt	gcc	gtt	atg	ttt	ggc	gat	gcc	gga	cac	ggg	gca	2044
Phe	Pro	Phe	Met	Phe	Ala	Val	Met	Phe	Gly	Asp	Ala	Gly	His	Gly	Ala	
	460					465				470						
ttg	atg	ctc	ata	gcc	gcg	ctc	tat	ttg	gtg	tta	aat	gaa	aag	aag	ttg	2092
Leu	Met	Leu	Ile	Ala	Ala	Leu	Tyr	Leu	Val	Leu	Asn	Glu	Lys	Lys	Leu	
475					480				485						490	
gga	gca	atg	aaa	cga	ggg	gag	att	ttc	gac	atg	gca	tac	act	gga	aga	2140
Gly	Ala	Met	Lys	Arg	Gly	Glu	Ile	Phe	Asp	Met	Ala	Tyr	Thr	Gly	Arg	
			495						500					505		
tat	gtt	att	cta	ctg	atg	gga	atc	ttt	tct	atc	tat	acc	ggg	ata	atg	2188
Tyr	Val	Ile	Leu	Leu	Met	Gly	Ile	Phe	Ser	Ile	Tyr	Thr	Gly	Ile	Met	
		510					515						520			
tac	aat	gat	att	ttt	tcc	aag	tcc	atg	cat	ttg	ttt	tcc	act	ggc	tgg	2236
Tyr	Asn	Asp	Ile	Phe	Ser	Lys	Ser	Met	His	Leu	Phe	Ser	Thr	Gly	Trp	
		525					530					535				
aaa	tgg	cct	tca	aac	ttt	caa	gag	ggc	gag	atg	att	gag	gct	caa	aaa	2284
Lys	Trp	Pro	Ser	Asn	Phe	Gln	Glu	Gly	Glu	Met	Ile	Glu	Ala	Gln	Lys	
	540					545					550					
gtt	ggg	gtt	tac	cca	ttt	gga	ttg	gac	tat	gcc	tgg	cac	ggg	tcg	gac	2332
Val	Gly	Val	Tyr	Pro	Phe	Gly	Leu	Asp	Tyr	Ala	Trp	His	Gly	Ser	Asp	
555					560				565						570	
aat	agt	ttg	tta	ttc	acc	aat	tca	tat	aaa	atg	aag	tta	tca	atc	ctt	2380
Asn	Ser	Leu	Leu	Phe	Thr	Asn	Ser	Tyr	Lys	Met	Lys	Leu	Ser	Ile	Leu	
			575						580					585		
cta	ggc	ttt	atc	cac	atg	tcg	tat	tca	tat	att	ttc	tca	tat	ctc	aac	2428
Leu	Gly	Phe	Ile	His	Met	Ser	Tyr	Ser	Tyr	Ile	Phe	Ser	Tyr	Leu	Asn	
		590					595						600			
tac	cac	tat	aaa	ggg	tcg	agg	atc	gat	att	gtc	ggc	aac	ttc	att	cct	2476
Tyr	His	Tyr	Lys	Gly	Ser	Arg	Ile	Asp	Ile	Val	Gly	Asn	Phe	Ile	Pro	
		605					610					615				
gga	tta	ata	ttt	atg	caa	tca	ata	ttt	ggg	tac	cta	tcg	tgg	gca	att	2524
Gly	Leu	Ile	Phe	Met	Gln	Ser	Ile	Phe	Gly	Tyr	Leu	Ser	Trp	Ala	Ile	
	620					625					630					
att	tac	aag	tgg	tca	aaa	gat	tgg	att	aaa	gat	gag	ctg	cct	gcg	cct	2572
Ile	Tyr	Lys	Trp	Ser	Lys	Asp	Trp	Ile	Lys	Asp	Glu	Leu	Pro	Ala	Pro	
635					640				645						650	
ggg	tta	tta	aac	atg	ttg	att	aac	atg	ttt	tta	tct	ccc	gga	gtt	gtc	2620
Gly	Leu	Leu	Asn	Met	Leu	Ile	Asn	Met	Phe	Leu	Ser	Pro	Gly	Val	Val	
			655					660						665		
gat	gaa	aaa	cta	tat	aca	ggc	cag	agc	ttt	ctt	caa	gtt	atc	ctt	ttg	2668
Asp	Glu	Lys	Leu	Tyr	Thr	Gly	Gln	Ser	Phe	Leu	Gln	Val	Ile	Leu	Leu	
		670					675						680			

cta gct gct cta gtt tgt gtt ccg tgg cta cta ctt tac aaa ccg ttg	2716
Leu Ala Ala Leu Val Cys Val Pro Trp Leu Leu Leu Tyr Lys Pro Leu	
685 690 695	
atg ttg aag cgc cag aac gat ata gct ctc agc aaa gga ttc aga agc	2764
Met Leu Lys Arg Gln Asn Asp Ile Ala Leu Ser Lys Gly Phe Arg Ser	
700 705 710	
ctc aga gac caa cgg gtg cat gag att ctt ctg gaa gca cag gaa aac	2812
Leu Arg Asp Gln Arg Val His Glu Ile Leu Leu Glu Ala Gln Glu Asn	
715 720 725 730	
gca ggc gaa gat atg ctg gtt gca gat tat gaa aat gaa gat gaa tcg	2860
Ala Gly Glu Asp Met Leu Val Ala Asp Tyr Glu Asn Glu Asp Glu Ser	
735 740 745	
tcg gag gag ttc aac ttt ggt gac gtt atg ata cat cag gtt atc cac	2908
Ser Glu Glu Phe Asn Phe Gly Asp Val Met Ile His Gln Val Ile His	
750 755 760	
acc att gaa ttt tgt ttg aat tgt ata tct cat act gca tcg tat ctc	2956
Thr Ile Glu Phe Cys Leu Asn Cys Ile Ser His Thr Ala Ser Tyr Leu	
765 770 775	
aga ttg tgg gcc tta tct ctt gcg cat gct caa ctc tcg act gtg ttg	3004
Arg Leu Trp Ala Leu Ser Leu Ala His Ala Gln Leu Ser Thr Val Leu	
780 785 790	
tgg tct atg acc att cag aat tcg ttc tcc gac tcc aac cct ggt tca	3052
Trp Ser Met Thr Ile Gln Asn Ser Phe Ser Asp Ser Asn Pro Gly Ser	
795 800 805 810	
ttt ttc tct gtc acc aag gtg gtg gtt ttg ttt gcc atg tgg ttt gtg	3100
Phe Phe Ser Val Thr Lys Val Val Val Leu Phe Ala Met Trp Phe Val	
815 820 825	
ttg act gtt tgc att tta gtc tta atg gaa gga acg tcg gct atg ttg	3148
Leu Thr Val Cys Ile Leu Val Leu Met Glu Gly Thr Ser Ala Met Leu	
830 835 840	
cac tcg tta aga ttg cat tgg gtg gaa gcg atg tcc aaa ttc ttc gag	3196
His Ser Leu Arg Leu His Trp Val Glu Ala Met Ser Lys Phe Phe Glu	
845 850 855	
ggc gaa ggc tat gct tac gaa cca ttt tcc ttc aag gct atc aac agc	3244
Gly Glu Gly Tyr Ala Tyr Glu Pro Phe Ser Phe Lys Ala Ile Asn Ser	
860 865 870	
gat gat gaa tagtatcgta tatattaataa ctagacagat ggaggtagtg	3293
Asp Asp Glu	
875	
tcacatgtgt caggtataga ttgccccgatg aataagcatc cttgaaattht agaaaaatca	3353
taggccttht gaatcgacat acaacgtcta aatatattta ctattcactt catagttcat	3413
cgtgatcgat gtccgaagtg gaactatctg cgctagaggt ggatgcatct gctggttggt	3473
gcgtagaggt cgcttttgct tcatctgcga agtcagtagt agatgaagag gctgcttctt	3533
tgcttttctc ttcttttctc tttgccgctc ttagttttct tgtgtataat gagtttagct	3593

cacgcaagcg gtatgcatgt gctcttttcgc ggaatttttg catttcctcg atggaattta 3653
 gatgaccttc ggatttcatg aacaatttta ctaactgttc gcggtcaatt gtatcaaag 3713
 catctggagt taatatgggc tcgagttcgc tcagtatctt accgagtgcg gccaatctg 3773
 actcagcaaa ggttgaattg tactgtaaat tctttggaac cttcagttta gtatattctc 3833
 tagtgacatt taggccaaga gagttaaaat tctcctcaag ttcattagct tgtgccactg 3893
 cagtatcatc aaatgcgta aaggtagaaa tggcttcaga aacatttttg cgtagaagct 3953
 caaatactgg gagatccaaa accttagcgt ctatggaaag gtaatggcta agcaacttat 4013
 acagactcat ggtttctggt aatctgtggt ggatatcgcc a 4054

<210> 22

<211> 877

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 31

<400> 22

Met Phe Leu Ser Met Arg Ala Glu Glu Ala Met Phe Arg Ser Ala Asp
 1 5 10 15

Met Thr Tyr Ile Glu Leu Tyr Ile Pro Leu Glu Ile Ala Arg Glu Val
 20 25 30

Val Cys Val Leu Gly Asn Leu Gly Ser Val Met Leu Lys Asp Met Asn
 35 40 45

Lys Asp Leu Ser Thr Phe Gln Arg Gly Tyr Val Asn Gln Val Arg Arg
 50 55 60

Phe Asp Glu Val Glu Arg Gln Val Gly Tyr Met Glu Gly Val Val Arg
 65 70 75 80

Arg His Lys Asn Glu Thr Trp Arg Tyr Leu Tyr Arg His Leu Gln Arg
 85 90 95

Glu Glu Gln Gln Glu Tyr Pro Gly Arg Glu His Pro Thr Leu Ala Gln
 100 105 110

Leu Ile Gly Ser Met His Thr His Ser Ile Asp Ser Val Asp Glu Val
 115 120 125

Ala Glu Glu Ile Met Gln Phe Glu Gly Arg Val Arg Gln Leu Asp Gln
 130 135 140

Ser Leu Val Ala Met Arg Glu Arg Leu Ser Lys Leu Val His Glu Arg
 145 150 155 160

Arg Val Met Phe Thr Cys Glu His Phe Leu Glu Val Asn Pro Gly Ile
 165 170 175

Gly Glu Arg Leu Pro Thr Arg Pro Ala Gly Phe Glu Ala Asp Glu Phe
 180 185 190

Glu Leu Thr Arg Val Gly Glu Glu Asp Glu Glu Thr Ala Ser Gln Leu
 195 200 205

Ser Phe Asp Ile Ser Asp Asp Ala Glu Thr Gln Leu Pro Gly Asp Met
 210 215 220

Arg Thr Leu Leu Glu Pro Val Tyr Arg His Gln Tyr Leu Leu Thr Gly
 225 230 235 240

Ser Ile Glu Arg Ala Lys Val Glu Ala Leu Asn Lys Ile Leu Trp Arg
 245 250 255

Leu Leu Arg Gly Asn Val Phe Phe Gln Asn Phe Pro Val Ser Val Ser
 260 265 270

Pro Val Glu Glu Asp Asp Thr Asp Leu Glu Thr Asp Cys Phe Ile Val
 275 280 285

Phe Thr His Gly Glu Val Leu Leu Ser Lys Ala Lys Lys Val Ile Glu
 290 295 300

Ser Leu Asn Gly Thr Ile Tyr Pro Phe Met Gln Asp Gly Ala Thr Val
 305 310 315 320

Gln Glu Leu Asn Asp Lys Ile Ala Asp Leu Lys Gln Ile Cys Ser Thr
 325 330 335

Thr Glu Gln Thr Leu His Thr Glu Leu Phe Leu Val Ala Asn Gln Leu
 340 345 350

Pro Met Trp Asn Ala Ile Ile Lys Arg Glu Lys Tyr Ile Tyr Ser Ala
 355 360 365

Leu Asn Leu Phe Arg Gln Glu Ser Gln Gly Leu Val Ala Glu Gly Trp
 370 375 380

Leu Pro Thr Tyr Asp Leu Pro Gly Val Gln Ala Ala Leu Lys Asp Tyr
 385 390 395 400

Gly Glu Ser Val Gly Ser Ala Asn Ser Ala Val Leu Asn Val Ile Ser
 405 410 415

Thr Thr Arg Thr Pro Pro Thr Phe His Arg Thr Asn Lys Phe Thr Gln
 420 425 430

Ala Phe Gln Ser Ile Val Asp Ala Tyr Gly Ile Ala Thr Tyr Lys Glu
 435 440 445

Val Asn Pro Gly Leu Ala Thr Ile Val Thr Phe Pro Phe Met Phe Ala
 450 455 460

Val Met Phe Gly Asp Ala Gly His Gly Ala Leu Met Leu Ile Ala Ala
 465 470 475 480

Leu Tyr Leu Val Leu Asn Glu Lys Lys Leu Gly Ala Met Lys Arg Gly
 485 490 495

Glu Ile Phe Asp Met Ala Tyr Thr Gly Arg Tyr Val Ile Leu Leu Met
 500 505 510

Gly Ile Phe Ser Ile Tyr Thr Gly Ile Met Tyr Asn Asp Ile Phe Ser
 515 520 525

Lys Ser Met His Leu Phe Ser Thr Gly Trp Lys Trp Pro Ser Asn Phe
 530 535 540

Gln Glu Gly Glu Met Ile Glu Ala Gln Lys Val Gly Val Tyr Pro Phe
 545 550 555 560

Gly Leu Asp Tyr Ala Trp His Gly Ser Asp Asn Ser Leu Leu Phe Thr
 565 570 575

Asn Ser Tyr Lys Met Lys Leu Ser Ile Leu Leu Gly Phe Ile His Met
 580 585 590

Ser Tyr Ser Tyr Ile Phe Ser Tyr Leu Asn Tyr His Tyr Lys Gly Ser

595	600	605
Arg Ile Asp Ile Val Gly Asn Phe Ile Pro Gly Leu Ile Phe Met Gln 610	615	620
Ser Ile Phe Gly Tyr Leu Ser Trp Ala Ile Ile Tyr Lys Trp Ser Lys 625	630	635 640
Asp Trp Ile Lys Asp Glu Leu Pro Ala Pro Gly Leu Leu Asn Met Leu 645	650	655
Ile Asn Met Phe Leu Ser Pro Gly Val Val Asp Glu Lys Leu Tyr Thr 660	665	670
Gly Gln Ser Phe Leu Gln Val Ile Leu Leu Leu Ala Ala Leu Val Cys 675	680	685
Val Pro Trp Leu Leu Leu Tyr Lys Pro Leu Met Leu Lys Arg Gln Asn 690	695	700
Asp Ile Ala Leu Ser Lys Gly Phe Arg Ser Leu Arg Asp Gln Arg Val 705	710	715 720
His Glu Ile Leu Leu Glu Ala Gln Glu Asn Ala Gly Glu Asp Met Leu 725	730	735
Val Ala Asp Tyr Glu Asn Glu Asp Glu Ser Ser Glu Glu Phe Asn Phe 740	745	750
Gly Asp Val Met Ile His Gln Val Ile His Thr Ile Glu Phe Cys Leu 755	760	765
Asn Cys Ile Ser His Thr Ala Ser Tyr Leu Arg Leu Trp Ala Leu Ser 770	775	780
Leu Ala His Ala Gln Leu Ser Thr Val Leu Trp Ser Met Thr Ile Gln 785	790	795 800
Asn Ser Phe Ser Asp Ser Asn Pro Gly Ser Phe Phe Ser Val Thr Lys 805	810	815
Val Val Val Leu Phe Ala Met Trp Phe Val Leu Thr Val Cys Ile Leu 820	825	830
Val Leu Met Glu Gly Thr Ser Ala Met Leu His Ser Leu Arg Leu His 835	840	845

Trp Val Glu Ala Met Ser Lys Phe Phe Glu Gly Glu Gly Tyr Ala Tyr
 850 855 860

Glu Pro Phe Ser Phe Lys Ala Ile Asn Ser Asp Asp Glu
 865 870 875

<210> 23

<211> 872

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 4

<400> 23

```

gatctcatgg aggattgatt atataactat ttcttgctga gaggaagtta tggctataat      60
cgccgtgtac acatgtatct ccaagacgct cactttttct tcatagaggg tgctccatat      120
aacctgatag agcttaagcg ggtctttgaa accaaagata aacttcttaa agtttattcc      180
atattactcg taaaatggaa ctctaaagtt ttacagcctc aagggtgagg cagtgcgcac      240
ttttttcatc cattccggca ccttgagaat atttgatgtt aagtgcctcg ttgttagtag      300
atagctttca tggctatctc gaatgtttga ctctgtattg gctccgcctg agtcaagctt      360
ttgaagctaa gaaaggacca acttgtaaaa cttcagacaa attttatcta ggaacatggc      420
ctgattcagc tcgaggtatc ctctttcaac tcaaatcaca ttaatgcact aaaactgagt      480
gtaagcaaga cctgcaaata ataaaatttg ggcacatttt cattaaattt ctcccatata      540
agatagtatc ttctcatgta aagcatatta atgaaaccca cacaaagacc agaagtatta      600
tactgaaaag tataggaaga aacagcaata atcatcccat tattaaataa ttctactaaa      660
ccggtcagga taaggtatat ttcagacatg agactgagct ttttccatag tccgaagaac      720
aatccttcac acgccacaat taaccagcac catttcatta taggatagaa atgttcagaa      780
attctctttg gatcaataga atctcacatg aaagaagtat tgacgtatct tctgattcct      840
gggcatgtaa acttaaaccg ctgttaagga tc                                     872

```

<210> 24

<211> 114

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 4

<400> 24

Pro Gln Arg Phe Lys Phe Thr Cys Pro Gly Ile Arg Lys Tyr Val Asn
1 5 10 15

Thr Ser Phe Met Asp Ser Ile Asp Pro Lys Arg Ile Ser Glu His Phe
20 25 30

Tyr Pro Ile Met Lys Trp Cys Trp Leu Ile Val Ala Cys Glu Gly Leu
35 40 45

Phe Phe Gly Leu Trp Lys Lys Leu Ser Leu Met Ser Glu Ile Tyr Leu
50 55 60

Ile Leu Thr Gly Leu Val Glu Leu Phe Asn Asn Gly Met Ile Ile Ala
65 70 75 80

Val Ser Ser Tyr Thr Phe Gln Tyr Asn Thr Ser Gly Leu Cys Val Gly
85 90 95

Phe Ile Asn Met Leu Tyr Met Arg Arg Tyr Tyr Leu Ile Trp Glu Lys
100 105 110

Phe Asn

<210> 25

<211> 2093

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 4

<400> 25

```

ccatgaatga gcgataatat atataacaag agtgggtgtag catggcatga gcctgcttgc      60
ggatggataa aggaggagtc ttcaggatag gaataatttc ccaagaatct attcatgaat      120
ggatccaca ataggtcttt aatgtgagtt acgtagagtc tcttatgaga tttttacaga      180
aatttgattt taaaaatggg catttagcat gtttagtcta tgtcgccagg ggaatgggtg      240
ttaaaaaattt tctcatttgt tgtgaattca agatatgcta tgatacacia ataaacgata      300
tacggactct accacatgca aattatagac taaaaagata cggttgataa aagcattttt      360
aaggagcaag attaaccgtc cagttgcgta atattatact taatttggtt tatagagggtg      420
tttttactcg tcttcatgat gtttatcggt attggataaa catgatggca ccactctcta      480
tcatatcgct cgcagaatag ccgatataat cttttcagta gcgaaaaaat attttgcggt      540
atcttaaata aatttaagct tagttaccat tttttttatt tagagaactg tctactgtga      600
tctcatggag gattgattat ataactatct cttgctgaga ggaagttatg gctataatcg      660
ccgtgtacac atgtattttcc aagacgctca ctttttcttc atagaggggtg ctccatataa      720
cctgatagag cttaagcggg tctttgaaac caaagataaa cttcttaaag tttattccat      780
attactcgta aaatggaact cttaaagtttt acagcctcaa gggtgaggca gtgacgactt      840
ttttcatcca ttccggcacc ttgagaatat ttgatgttaa gtgctctggt gttagtagat      900
agctttcatg gctatctcga atgtttgact ctgtattggc tccgcctgag tcaagctttt      960
gaagctaaga aaggaccaac ttgtaaaact tcagacaaat tttatctagg aacatggcct     1020
gattcagctc gaggtatcct ctttcaactc aaatcacatt aatgcactaa aactgagtgt     1080
aagcaagacc tgcaaataat aaaatttggg cacattttca ttaaatttct cccatataag     1140
atagtatctt ctcatgtaaa gcatattaat gaaaccaca caaagaccag aagtattata     1200
ctgaaaagta taggaagaaa cagcaataat catccatta ttaaataatt ctactaaacc     1260
ggtcaggata aggtatatct cagacatgag actgagcttt ttccatagtc cgaagaacaa     1320
tccttcacac gccacaatta accagcacca tttcattata ggatagaaat gttcagaaat     1380
tctcttttga tcaatagaat ctccatgaa agaagtattg acgtattttc tgattcctgg     1440
gcatgtaaac ttaaaccgct gttaaggatc gcaatagtct tcaatattat gaagttgccca     1500
attaataacg ccaagattca caaagatctg tataagaatc attaaaagct gaccccgaaa     1560
tattggaacg ggtgggtatct ttgcgtaatg tgccaactta atatcgccaa gatagttatc     1620
agtttgata tcaatattat acgcgaatga cttgataaaa gttagcgcat aatgatttct     1680

```

cggcactttt caaaccactt attacctgca acagcagcat tggtgctata gtaatactat 1740
ctgtgacttg tagcaacgcg acggaaagtg atagatcgga gttctaactct atggaaaacc 1800
atattgctca gatcaattcg acttccaagg ctatactacc agagatattg caaagggttg 1860
gaagaacggc ccctacagtc actcctgctg ccattgggtc ttaagagata ccatatctta 1920
aatacccctt gtcgggcaat gatgaggatt agcaacattc attgatagtt ggtatccttt 1980
tcttagtacc aatgcggtgt gcagggtgta tgggtttacag gcggaacaat ggaggatatg 2040
cagtttctgg acaaggataa tcaaggaatg aaaccattag gagtggggcca ata 2093

<210> 26

<211> 2093

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (738) .. (1037)

<223>

<220>

<221> misc_feature

<223> Oligo 4

<400> 26

tattggccca ctctaatgg tttcattcct tgattatcct tgtccagaaa ctgcatatcc 60
tccattgttc cgctgtaaa ccattacacc tgcacaccgc attggtacta agaaaaggat 120
accaactatc aatgaatggt gctaattcct atcattgccc gacaaggggt atttaagata 180
tggtatctct taagaaccaa tggcagcagg agtgactgta gggaccgttc ttcacaacct 240
ttgcaatatc tctggtagta tagccttggg agtcgaattg atctgagcaa tatgggttttc 300
catagattag aactccgata tatcactttc cgtcgcgttg ctacaagtca cagatagtat 360
tactatagca acaatgctgc tggtgcaggc aataagtggc ttgaaaagtg ccgagaaatc 420
attatgcgct aactttttatc aagtcattcg cgtataatat tgatatccaa actgataaact 480

atcttggcga tattaagttg gcacattacg caaagatacc acccgttcca atatttcggg	540
gtcagctttt aatgattcctt atacagatct ttgtgaatct tggcggttatt aattggcaac	600
ttcataatat tgagaactat tgcgatcctt aacagcgggt taagtttaca tgcccaggaa	660
tcagaaaata cgtcaatact tctttcatgt gagattctat tgatccaaag agaatttctg	720
aacattttcta tcctata atg aaa tgg tgc tgg tta att gtg gcg tgt gaa	770
Met Lys Trp Cys Trp Leu Ile Val Ala Cys Glu	
1 5 10	
gga ttg ttc ttc gga cta tgg aaa aag ctc agt ctc atg tct gaa ata	818
Gly Leu Phe Phe Gly Leu Trp Lys Lys Leu Ser Leu Met Ser Glu Ile	
15 20 25	
tac ctt atc ctg acc ggt tta gta gaa tta ttt aat aat ggg atg att	866
Tyr Leu Ile Leu Thr Gly Leu Val Glu Leu Phe Asn Asn Gly Met Ile	
30 35 40	
att gct gtt tct tcc tat act ttt cag tat aat act tct ggt ctt tgt	914
Ile Ala Val Ser Ser Tyr Thr Phe Gln Tyr Asn Thr Ser Gly Leu Cys	
45 50 55	
gtg ggt ttc att aat atg ctt tac atg aga aga tac tat ctt ata tgg	962
Val Gly Phe Ile Asn Met Leu Tyr Met Arg Arg Tyr Tyr Leu Ile Trp	
60 65 70 75	
gag aaa ttt aat gaa aat gtg ccc aaa ttt tat tat ttg cag gtc ttg	1010
Glu Lys Phe Asn Glu Asn Val Pro Lys Phe Tyr Tyr Leu Gln Val Leu	
80 85 90	
ctt aca ctc agt ttt agt gca tta atg tgatttgagt tgaaagagga	1057
Leu Thr Leu Ser Phe Ser Ala Leu Met	
95 100	
tacctcgagc tgaatcaggc catgttccta gataaaattht gtctgaagtt ttacaagttg	1117
gtccttttctt agcttcaaaa gcttgactca ggcggagcca atacagagtc aaacattcga	1177
gatagccatg aaagctatct actaacaaca gagcacttaa catcaaatat tctcaaggtg	1237
ccggaatgga tgaaaaaagt cgtcactgcc tcacccttga ggctgtaaaa ctttagagtt	1297
ccatttttacg agtaatatgg aataaacttt aagaagttha tctttggttt caaagaccg	1357
cttaagctct atcaggttat atggagcacc ctctatgaag aaaaagtgag cgtcttgga	1417
atacatgtgt acacggcgat tatagccata acttcctctc agcaagaaat agttatataa	1477
tcaatcctcc atgagatcac agtagacagt tctctaaata aaaaaaatgg taactaagct	1537
taaattttatt taagataacg caaaatattt tttcgctact gaaaagatta tatcggctat	1597
tctgcgagcg atatgataga gagtgggtgcc atcatgttta tccaataccg ataaacatca	1657
tgaagacgag taaaaacacc tctataaacc aaattaagta taatattacg caactggacg	1717
gttaatcttg ctctttaaaa atgcttttat caaccgtatc tttttagtct ataatttgca	1777
tgtggtagag tccgtatatc gtttatttgt gtatcatagc atatcttgaa ttcacaacaa	1837

```

atgagaaaat ttttaaaccac cattcccctg ggcacataga ctaaaccatgc taaatgacca 1897
tttttaaaat caaatctctg taaaaatctc ataagagact ctacgtaact cacattaaag 1957
acctattgtg gataccattc atgaatagat tcttgggaaa ttattcctat cctgaagact 2017
cctcctttat ccatccgcaa gcaggctcat gccatgctac accactcttg ttatatatat 2077
tatcgctcat tcatgg 2093

```

<210> 27

<211> 100

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 4

<400> 27

```

Met Lys Trp Cys Trp Leu Ile Val Ala Cys Glu Gly Leu Phe Phe Gly
1           5           10          15

```

```

Leu Trp Lys Lys Leu Ser Leu Met Ser Glu Ile Tyr Leu Ile Leu Thr
20          25          30

```

```

Gly Leu Val Glu Leu Phe Asn Asn Gly Met Ile Ile Ala Val Ser Ser
35          40          45

```

```

Tyr Thr Phe Gln Tyr Asn Thr Ser Gly Leu Cys Val Gly Phe Ile Asn
50          55          60

```

```

Met Leu Tyr Met Arg Arg Tyr Tyr Leu Ile Trp Glu Lys Phe Asn Glu
65          70          75          80

```

```

Asn Val Pro Lys Phe Tyr Tyr Leu Gln Val Leu Leu Thr Leu Ser Phe
85          90          95

```

```

Ser Ala Leu Met
100

```

<210> 28

<211> 853

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 6

<400> 28

```
gatcatattc cccttgtagg tctagccgca cgggtaaacg gcaatgtatg tcgcagaaca      60
agtccccaca cctccggcag ttctggattc ctgtcctcgg tgtgagcttt cgcccgcatac      120
tcgagcattg tgacttccct ggcacaagag gtttgtaatg tgacctgttt ctgccggatg      180
tctcccgact ggacgccaga cccaccgaag ttggtgtagt acttgcaatg ttcacsggaa      240
ccgctgtctg tggatagcga tcgttcatct ttaaaattgt gttcaacatac cagatgcata      300
ttcaactgtt ccaagttttc aatatttcga ttgcatacgg ggcactcgac cgattccgag      360
gcgtcagagg cagcagcatt actgtactct gtaggtatgt gttcatctgc ctccgggggtt      420
ggcattactg ggccacctag ttggaactga catttatgat ataacgcaca gcactaatgt      480
gtgaaacgca aaagttcacc aactgtcaa aaaccaccac acagtcaact cgagatagac      540
acaataatga tggaagagct gtttatgcta ccagttattg gtatcacggc gttgttggtg      600
tataaatacg cttacgatat cctttggtac aagttacaga atctatttga actgattcct      660
tcatcaagct catctcacc ctactcttcg gctatcgcta gcattaataa gtctcagaat      720
gggtttctgt acaagctata cacggaatat tcagtttctt ctaacaatgt gctgcgagtg      780
attcggactg ctagtctccg ggactgttgg ctgctgtgcg tctgtggctg tggagatagt      840
gctatggcag atc                                                    853
```

<210> 29

<211> 38

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 6

<400> 29

Glu Ser Val Glu Cys Pro Val Cys Asn Arg Asn Ile Glu Asn Leu Glu
1 5 10 15

Gln Leu Asn Met His Leu Asp Val Glu His Asn Phe Lys Asp Glu Arg
20 25 30

Ser Leu Ser Thr Asp Ser
35

<210> 30

<211> 55

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 6

<400> 30

Arg Ser His Tyr Lys Pro Leu Val Pro Gly Lys Ser Gln Cys Ser Arg
1 5 10 15

Cys Gly Arg Lys Leu Thr Pro Arg Thr Gly Ile Gln Asn Cys Arg Arg
20 25 30

Cys Gly Asp Leu Phe Cys Asp Ile His Cys Arg Leu Pro Val Arg Leu
35 40 45

Asp Leu Gln Gly Glu Tyr Asp
50 55

<210> 31

<211> 2389

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (428) .. (1993)

<223>

<220>

<221> misc_feature

<223> Oligo 6

<400> 31

gatctgccat agcactatct ccacagccac agacgcacag cagccaacag tcccggagac	60
tagcgtccga atcactcgca gcacattggt agaagaaact gaatattccg tgtatagctt	120
gtacagaaac ccattctgag acttattaat gctagcgata gccgaagagt aggggtgaga	180
tgagcttgat gaaggaatca gttcaaatag attctgtaac ttgtacaaa ggatatcgta	240
agcgtattta tacaccaaca acgccgtgat accaataact ggtagcataa acagctcttc	300
catcattatt gtgtctatct cgagttgact gtgtggtggt ttttgacagt gtggtgaact	360
tttgcgtttc acacattagt gctgtgcgtt atatcataaa tgtcagttcc aactaggtgg	420
cccagta atg cca acc ccg gag gca gat gaa caa ata cct aca gag tac	469
Met Pro Thr Pro Glu Ala Asp Glu Gln Ile Pro Thr Glu Tyr	
1 5 10	
agt aat gct gct gcc tct gac gcc tcg gaa tcg gtc gag tgc ccc gta	517
Ser Asn Ala Ala Ala Ser Asp Ala Ser Glu Ser Val Glu Cys Pro Val	
15 20 25 30	
tgc aat cga aat att gaa aac ttg gaa cag ttg aat atg cat ctg gat	565
Cys Asn Arg Asn Ile Glu Asn Leu Glu Gln Leu Asn Met His Leu Asp	
35 40 45	
gtt gaa cac aat ttt aaa gat gaa cga tcg cta tcc aca gac agc ggt	613
Val Glu His Asn Phe Lys Asp Glu Arg Ser Leu Ser Thr Asp Ser Gly	
50 55 60	
tcc gtg aac att gca agt act aca cca act tcg gtg ggt ctg gcg tcc	661
Ser Val Asn Ile Ala Ser Thr Pro Thr Ser Val Gly Leu Ala Ser	
65 70 75	
agt cgg gag aca tcc ggc aga aac agg tca cat tac aaa cct ctt gtg	709
Ser Arg Glu Thr Ser Gly Arg Asn Arg Ser His Tyr Lys Pro Leu Val	
80 85 90	
cca ggg aag tca caa tgc tcg aga tgc ggg cga aag ctc aca ccg agg	757
Pro Gly Lys Ser Gln Cys Ser Arg Cys Gly Arg Lys Leu Thr Pro Arg	
95 100 105 110	

aca gga atc cag aac tgc cgg agg tgt ggg gac ttg ttc tgc gac ata	805
Thr Gly Ile Gln Asn Cys Arg Arg Cys Gly Asp Leu Phe Cys Asp Ile	
115 120 125	
cat tgc cgt tta ccc gtg cgg cta gac cta caa ggg gaa tat gat ccg	853
His Cys Arg Leu Pro Val Arg Leu Asp Leu Gln Gly Glu Tyr Asp Pro	
130 135 140	
aga aac ggc gac tgg tgc aag tgc tgc cat gcc tgc atg gcg ggg cgc	901
Arg Asn Gly Asp Trp Cys Lys Cys Cys His Ala Cys Met Ala Gly Arg	
145 150 155	
cct gga tac aac aaa ctg ggc tta tgc gtg gac cgg acc gac gag ttt	949
Pro Gly Tyr Asn Lys Leu Gly Leu Ser Val Asp Arg Thr Asp Glu Phe	
160 165 170	
ata cgc cat cgt acc tct aaa aac gag gat aaa cag ctt cgc att ctg	997
Ile Arg His Arg Thr Ser Lys Asn Glu Asp Lys Gln Leu Arg Ile Leu	
175 180 185 190	
cag ctt gag aat cgg ctt gtg cgt ctt gtt gac ggc atc gct act att	1045
Gln Leu Glu Asn Arg Leu Val Arg Leu Val Asp Gly Ile Ala Thr Ile	
195 200 205	
gtt cgc gcg cac aac cag tct ttg ttt tat gga agt gga atg tat agg	1093
Val Arg Ala His Asn Gln Ser Leu Phe Tyr Gly Ser Gly Met Tyr Arg	
210 215 220	
gaa atc act gct cta cag aag tcc gtt acg cca tgg aaa gaa aat tca	1141
Glu Ile Thr Ala Leu Gln Lys Ser Val Thr Pro Trp Lys Glu Asn Ser	
225 230 235	
cag gca tca agc tgc tat ctc tgt tct cgg ccg ttc aat ctt ctt ctt	1189
Gln Ala Ser Ser Cys Tyr Leu Cys Ser Arg Pro Phe Asn Leu Leu Leu	
240 245 250	
cgc aag cac cat tgc aag ttg tgt ggt ctg ata gtc tgc gaa aac aac	1237
Arg Lys His His Cys Lys Leu Cys Gly Leu Ile Val Cys Glu Asn Asn	
255 260 265 270	
ttt acc aac tgt tcc aag gag ttc ccg atc gcg cag ctg gtg agt gct	1285
Phe Thr Asn Cys Ser Lys Glu Phe Pro Ile Ala Gln Leu Val Ser Ala	
275 280 285	
gcc aca gac cta cca ttc agg agc aat ccc cag gaa tta gcc gca ttg	1333
Ala Thr Asp Leu Pro Phe Arg Ser Asn Pro Gln Glu Leu Ala Ala Leu	
290 295 300	
ccg gtt cga cta cgc gtc tgc gtg gtt tgc ata aga tcc gtg ttt ctg	1381
Pro Val Arg Leu Arg Val Cys Val Val Cys Ile Arg Ser Val Phe Leu	
305 310 315	
cgc gca agg ttg cag gac aat ctc gca aac gat gct tct cag cta ttt	1429
Arg Ala Arg Leu Gln Asp Asn Leu Ala Asn Asp Ala Ser Gln Leu Phe	
320 325 330	
tcc aag tac act gag ttg cag cgt gta tca aga gcc atc ttg cgt ata	1477
Ser Lys Tyr Thr Glu Leu Gln Arg Val Ser Arg Ala Ile Leu Arg Ile	
335 340 345 350	

atg cct cgt ttc gag caa ctg ctc ggg gac ctc aac gca ccc gac gcg	1525
Met Pro Arg Phe Glu Gln Leu Leu Gly Asp Leu Asn Ala Pro Asp Ala	
355 360 365	
agc ccg aat cga agc gag cta gac gag ctg gcg cac ctg cga agg aaa	1573
Ser Pro Asn Arg Ser Glu Leu Asp Glu Leu Ala His Leu Arg Arg Lys	
370 375 380	
ctc ttg gag act ttc aag ctg tac gac acc ata gcc aag cag ata ttt	1621
Leu Leu Glu Thr Phe Lys Leu Tyr Asp Thr Ile Ala Lys Gln Ile Phe	
385 390 395	
gcc atc gac ccg gcc aat acc gca gaa ttg aag atc cag cag gct ata	1669
Ala Ile Asp Pro Ala Asn Thr Ala Glu Leu Lys Ile Gln Gln Ala Ile	
400 405 410	
aag gcc aaa tcg atg tcc ttc ata cag gat aaa atg cta ccg cta aag	1717
Lys Ala Lys Ser Met Ser Phe Ile Gln Asp Lys Met Leu Pro Leu Lys	
415 420 425 430	
aat atc ccc ggc ctg ctg aag ccg aag gat gcc gaa cct gag atc aat	1765
Asn Ile Pro Gly Leu Leu Lys Pro Lys Asp Ala Glu Pro Glu Ile Asn	
435 440 445	
tac act act agc aat cta ctt ttc aac aat ctg act gtc cgc gag gtt	1813
Tyr Thr Thr Ser Asn Leu Leu Phe Asn Asn Leu Thr Val Arg Glu Val	
450 455 460	
aaa ctc tac cgt gaa cag ctg atg gta ctc aaa gag cag agg ttt ata	1861
Lys Leu Tyr Arg Glu Gln Leu Met Val Leu Lys Glu Gln Arg Phe Ile	
465 470 475	
gtg gag ggc atg ctc gag aac gcc aag aaa cag cgg cgt ttt gaa gag	1909
Val Glu Gly Met Leu Glu Asn Ala Lys Lys Gln Arg Arg Phe Glu Glu	
480 485 490	
gtt aat acg tta aag gaa aat acc aaa gag cta gac aat cag ata gcc	1957
Val Asn Thr Leu Lys Glu Asn Thr Lys Glu Leu Asp Asn Gln Ile Ala	
495 500 505 510	
cag ctc gaa gaa acc cta ggc gac cag ggt ttt gtt tagtatctag	2003
Gln Leu Glu Glu Thr Leu Gly Asp Gln Gly Phe Val	
515 520	
catggagttt tttgcttaac tataattact gtgtagatgc cgcagatagc atgtcgtagc	2063
ataattgcga attttcacca acatgaaaa gtgtatgtgt ataaggcatc cagtgaactc	2123
ctaacatgct gatgagggtt taagtaaaga tatcactagc aatgaacgta agtgcagttt	2183
ttgagcttta tgtcctctgt agaacataat attaacgaca gggggatagg atgaaagaag	2243
acagcagtta tttgagctga acagtgaagc ctggtctgga attgatgcgt tcccgaataa	2303
aaccagcaag cttgactcaa gcatcaagag aaacacaggg tttatcaaaa agctgaaaca	2363
gggtatcacg aaagactcga aagatc	2389

<211> 522

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 6

<400> 32

Met Pro Thr Pro Glu Ala Asp Glu Gln Ile Pro Thr Glu Tyr Ser Asn
 1 5 10 15

Ala Ala Ala Ser Asp Ala Ser Glu Ser Val Glu Cys Pro Val Cys Asn
 20 25 30

Arg Asn Ile Glu Asn Leu Glu Gln Leu Asn Met His Leu Asp Val Glu
 35 40 45

His Asn Phe Lys Asp Glu Arg Ser Leu Ser Thr Asp Ser Gly Ser Val
 50 55 60

Asn Ile Ala Ser Thr Thr Pro Thr Ser Val Gly Leu Ala Ser Ser Arg
 65 70 75 80

Glu Thr Ser Gly Arg Asn Arg Ser His Tyr Lys Pro Leu Val Pro Gly
 85 90 95

Lys Ser Gln Cys Ser Arg Cys Gly Arg Lys Leu Thr Pro Arg Thr Gly
 100 105 110

Ile Gln Asn Cys Arg Arg Cys Gly Asp Leu Phe Cys Asp Ile His Cys
 115 120 125

Arg Leu Pro Val Arg Leu Asp Leu Gln Gly Glu Tyr Asp Pro Arg Asn
 130 135 140

Gly Asp Trp Cys Lys Cys Cys His Ala Cys Met Ala Gly Arg Pro Gly
 145 150 155 160

Tyr Asn Lys Leu Gly Leu Ser Val Asp Arg Thr Asp Glu Phe Ile Arg
 165 170 175

His Arg Thr Ser Lys Asn Glu Asp Lys Gln Leu Arg Ile Leu Gln Leu
 180 185 190

Glu Asn Arg Leu Val Arg Leu Val Asp Gly Ile Ala Thr Ile Val Arg
 195 200 205

Ala His Asn Gln Ser Leu Phe Tyr Gly Ser Gly Met Tyr Arg Glu Ile
 210 215 220

Thr Ala Leu Gln Lys Ser Val Thr Pro Trp Lys Glu Asn Ser Gln Ala
 225 230 235 240

Ser Ser Cys Tyr Leu Cys Ser Arg Pro Phe Asn Leu Leu Leu Arg Lys
 245 250 255

His His Cys Lys Leu Cys Gly Leu Ile Val Cys Glu Asn Asn Phe Thr
 260 265 270

Asn Cys Ser Lys Glu Phe Pro Ile Ala Gln Leu Val Ser Ala Ala Thr
 275 280 285

Asp Leu Pro Phe Arg Ser Asn Pro Gln Glu Leu Ala Ala Leu Pro Val
 290 295 300

Arg Leu Arg Val Cys Val Val Cys Ile Arg Ser Val Phe Leu Arg Ala
 305 310 315 320

Arg Leu Gln Asp Asn Leu Ala Asn Asp Ala Ser Gln Leu Phe Ser Lys
 325 330 335

Tyr Thr Glu Leu Gln Arg Val Ser Arg Ala Ile Leu Arg Ile Met Pro
 340 345 350

Arg Phe Glu Gln Leu Leu Gly Asp Leu Asn Ala Pro Asp Ala Ser Pro
 355 360 365

Asn Arg Ser Glu Leu Asp Glu Leu Ala His Leu Arg Arg Lys Leu Leu
 370 375 380

Glu Thr Phe Lys Leu Tyr Asp Thr Ile Ala Lys Gln Ile Phe Ala Ile
 385 390 395 400

Asp Pro Ala Asn Thr Ala Glu Leu Lys Ile Gln Gln Ala Ile Lys Ala
 405 410 415

Lys Ser Met Ser Phe Ile Gln Asp Lys Met Leu Pro Leu Lys Asn Ile
 420 425 430

Pro Gly Leu Leu Lys Pro Lys Asp Ala Glu Pro Glu Ile Asn Tyr Thr
 435 440 445

Thr Ser Asn Leu Leu Phe Asn Asn Leu Thr Val Arg Glu Val Lys Leu
 450 455 460

Tyr Arg Glu Gln Leu Met Val Leu Lys Glu Gln Arg Phe Ile Val Glu
 465 470 475 480

Gly Met Leu Glu Asn Ala Lys Lys Gln Arg Arg Phe Glu Glu Val Asn
 485 490 495

Thr Leu Lys Glu Asn Thr Lys Glu Leu Asp Asn Gln Ile Ala Gln Leu
 500 505 510

Glu Glu Thr Leu Gly Asp Gln Gly Phe Val
 515 520

<210> 33

<211> 975

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 146

<400> 33
 gatccggacg acagggctgc ggcacaggag tcgtaccgta ggggcggggc agtccgaggc 60
 cagcacgaat gggttccggc cgttcacgga cgagggcatg ttcttccgag gggccggggg 120
 ggctccggag gacctgtttg actttttctt tcggggcggg ggccccggcg gcccgttcgg 180
 gatggctgat ccctatgatt cgtttgggcc ctttgggggc gctacgacgt ttacgtttgg 240
 cgggccccgc ggcttcaagg tgtacagcgg agggctctggg gggcagttcc gccgcggacc 300
 gtttggcatg gcacaggctg ccaccgaagc gcaacgcaa cgtgcgaacg gccagccgga 360
 ggcgcaagat ccgctacaac acgtagtgtt cgttctcctg atagtgttct tcttctctgt 420
 actgccaagt ctgggcttct gacttggagc agtatccgtt gtaccatata gggccggcgc 480
 cgacctgacc aaggcttccc atggctggcc ttccatgatg gcatttcgta aatacctatt 540
 tttatagtca acaagtataa ataaataata atatacaaat aatttacaaa tctgcagcac 600

```

gtcgcgtgac gctgactgcg acataaaaat gttcaagata gtttgattag aatgcacaaa    660
caaacattgc aacacttaca taaaccagat gaagctacag ccgctattcc aggttttgca    720
ttctgcgttc tattatttcg tcaatatacct cggctgctgg cttaagcttt agcttcctgc    780
ctaatttggc ggcaaatacca tcgtttcccg agccgccaga attgtcaggt actagatttt    840
ttctcttgat aagtttgcca ctcggtaact gtcataattc ctctgtgtta tatgtggcac    900
tgctgtgtat agtcactttg ctaggcgttc caggaggtaa ctcggtcgca ccgggtagtg    960
tattcttcct ggatc                                                    975

```

<210> 34

<211> 66

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 146

<400> 34

```

Ser Ser Ala Thr Tyr Asn Thr Glu Glu Tyr Glu Gln Leu Pro Ser Gly
1           5           10           15

```

```

Lys Leu Ile Lys Arg Lys Asn Leu Val Pro Asp Asn Ser Gly Gly Ser
          20          25          30

```

```

Gly Asn Asp Gly Phe Ala Ala Lys Leu Gly Arg Lys Leu Lys Leu Lys
          35          40          45

```

```

Pro Ala Ala Glu Asp Ile Asp Glu Ile Ile Glu Arg Arg Met Gln Asn
          50          55          60

```

```

Leu Glu
65

```

<210> 35

<211> 2509

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (882) .. (1208)

<223>

<220>

<221> misc_feature

<223> Oligo 146

<400> 35
ctgcagtttc tgaaggtgct ggtggcagat ccctacggtt tgaacaagat ccgcaccgcg 60
ttcgagaagg gctacatcac gccgaatgac acatgggtcc aggggtggcac gacgacatac 120
gacaacattg cctatacgct gcgcctgatg atcgactgcg gcattgttgg catgtcgtgg 180
atcacactac caaagggcaa gtatgccatg atcccaaaga acaaaaaaat atctacgtgc 240
cagctggaag tgtcaataaa ctataaggac ctgatttcgc ggccggcaga cagtgcactgg 300
tcgcacagtg ctccattgcg aatcctctct ttcgacatag aatgtgctgg ccgcgtgggc 360
gtgtttccag aaccggaggt tgatcctgtc attcagatag cgaatgttgt cagtattgcg 420
ggagagtcaa aaccatttat tcgcaatgtc ttcacagttg acacctgtgc gccattact 480
ggctcgcaaa tttttgagca tgagacagaa gaagcgatgc ttaaacactg gcgtgacttc 540
cttgttgagg ttgacccga cgttatcatt ggttacaata cgacaaattt tgatcttcct 600
tacctgatca acagagctgc agccttgaac gtgtcttcac ttcctactt tggccgtttg 660
gttaactcca agcaggtcat caaagaaact gtattttcat ccaaggctca tggaacgcgg 720
gagtcgaagt cgatcaatat cgagggccgc cttcagttgg atatttttca gtttgtaaga 780
cgagaatatc agttgagatc gtacactttg aattcgggtgc gtttaatgag ttgaagcaag 840
gatggacatg gcaacgtgat cctccaacaa ttaagcgctt g atg aac aag gcg att 896
Met Asn Lys Ala Ile
1 5
gga aat aaa agc gat act. ttt gat ggc gtt atc aac cat acg cga tcc 944
Gly Asn Lys Ser Asp Thr Phe Asp Gly Val Ile Asn His Thr Arg Ser
10 15 20
agg aag aat aca cta ccc ggt gcg acc gag tta cct cct gga acg cct 992
Arg Lys Asn Thr Leu Pro Gly Ala Thr Glu Leu Pro Pro Gly Thr Pro
25 30 35

agc aaa gtg act ata cac agc agt gcc aca tat aac aca gag gaa tat	1040
Ser Lys Val Thr Ile His Ser Ser Ala Thr Tyr Asn Thr Glu Glu Tyr	
40 45 50	
gag cag tta ccg agt ggc aaa ctt atc aag aga aaa aat cta gta cct	1088
Glu Gln Leu Pro Ser Gly Lys Leu Ile Lys Arg Lys Asn Leu Val Pro	
55 60 65	
gac aat tct ggc ggc tcg gga aac gat gga ttt gcc gcc aaa tta ggc	1136
Asp Asn Ser Gly Gly Ser Gly Asn Asp Gly Phe Ala Ala Lys Leu Gly	
70 75 80 85	
agg aag cta aag ctt aag cca gca gcc gag gat att gac gaa ata ata	1184
Arg Lys Leu Lys Leu Lys Pro Ala Ala Glu Asp Ile Asp Glu Ile Ile	
90 95 100	
gaa cgc aga atg caa aac ctg gaa tagcggctgt agcttcatct ggtttatgta	1238
Glu Arg Arg Met Gln Asn. Leu Glu	
105	
agtgttgcaa tgtttgtttg tgcattctaa tcaaaactatc ttgaacattt ttatgtcgca	1298
gtcagcgtca cgcgacgtgc tgcagatttg taaattattt gtatattatt atttatttat	1358
acttgttgac tataaaaaata ggtatttacg aaatgccatc atggaaggcc agccatggga	1418
agccttggtc aggtcggcgc cggccctgta tggtaacaacg gatactgctc caagtcagaa	1478
gcccagactt ggcagtagca ggaagagaag cactatcagg agaacgaaca ctacgtgttg	1538
tagcggatct tgcgcctccg gctggccgtt cgcacgttgg cgttgcgctt cgggtggcagc	1598
ctgtgccatg ccaaacggtc cgcggcgga ctgccccca gaccaccgc tgtacacctt	1658
gaagcccgcg gggccgcaa acgtaaacgt cgtagcgccc ccaaagggcc caaacgaatc	1718
atagggatca gccatcccga acggggcgcc ggggcccccg ccccgaaaga aaaagtcaaa	1778
caggtcctcc ggagcccccc cggcccctcg gaagaacatg ccctcgtccg tgaacggccg	1838
gaacccattc gtgctggcct cggactgcc cgcccctacg gtacgactcc tgtgccgcag	1898
ccctgtcgtc cggatcgtag ccagctgggt cgtacacgcg gcgcttcttc tcgtcagaca	1958
aaacctcgaa cgcacgattt accttcttga acgcttccgc cgccctcggg tgccgatttt	2018
tgtccggatg tagtttgatt gccatcttgc ggtacgcctt ctttatgtcc ccatcgctcg	2078
ccttctcgtc cacctgtagc agctcataga acgaatgctt gtcccttatcc acaatcagca	2138
gcgtcagctt ttcttgctct tctgtgtact ctttctccgc catcttgata gctatcttag	2198
ctctccgcca atctgccaga ccttgttctt cttgccaccg agctacaatt cctcgagcct	2258
ctcaacccaaa ctaccttata acagctcctt cagatcacac gctgctggaa tcgatgactg	2318
gtccccggtc tgtcactatt catgcacaaa gcaactatag aacactcctc ttttcttaaa	2378
tgtaaaagtt gaagccaccg tgcagcgccg ttacgggaa gttccccaag gtagcacaac	2438
agaaggtaaa gatgaaacga ctacgcaccc gcataccggc ctaggggcaa cggcgacatt	2498

cggcgtacag c

2509

<210> 36

<211> 109

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 146

<400> 36

Met Asn Lys Ala Ile Gly Asn Lys Ser Asp Thr Phe Asp Gly Val Ile
1 5 10 15

Asn His Thr Arg Ser Arg Lys Asn Thr Leu Pro Gly Ala Thr Glu Leu
20 25 30

Pro Pro Gly Thr Pro Ser Lys Val Thr Ile His Ser Ser Ala Thr Tyr
35 40 45

Asn Thr Glu Glu Tyr Glu Gln Leu Pro Ser Gly Lys Leu Ile Lys Arg
50 55 60

Lys Asn Leu Val Pro Asp Asn Ser Gly Gly Ser Gly Asn Asp Gly Phe
65 70 75 80

Ala Ala Lys Leu Gly Arg Lys Leu Lys Leu Lys Pro Ala Ala Glu Asp
85 90 95

Ile Asp Glu Ile Ile Glu Arg Arg Met Gln Asn Leu Glu
100 105

<210> 37

<211> 1031

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 56

<400> 37

```
gatccactaa taattggacg atacgcgcgt acccttcagt cgctgccgca aagattgggg      60
tccagccaaa aagcttctca ccaagttcca tatttgcacc ctgtttaact agatattcgg      120
ccgcctcata gatgtctaac ttgcaagcaa cataaagagc agtctcgaga ttttccggct      180
cctgatagtc aatatcaaag cccttagcag acaataaaga atttaacaaa gacggcgagt      240
ttagtcttgt cgcgaggtgt agcaaacgag ggctgttaag agatttctct gggttcataa      300
acgacaatag cgtgctgaca gtaagagggg gtgtcccgat aacagccaaa tgcaatgggtg      360
tgaggttctc cgaatctccc cacacgtcaa tatcatcgat tgcaacgtct gcattccaag      420
catcccactc actcagagcy tgcagaatta tcctgggtamc ctcacataga ccatattggg      480
cagagtaatg tagcggcggt cgcttgtagt tatctctttg taacagcgac ggtcttaaatt      540
gggctggcaa ctgttgtaat atgtgtgaca ggggagctgg agagtcgtca gagttaacgc      600
catccggacc aaaagcacca acaagtctag tattgggttc cgggggtatc gctgcctcta      660
gatcgagaga gtcgctcagt agcgagttta tggtatcacg ctcttcgaga ctcttggtcc      720
gctttttgcc cagtgcattt acatgatgat ggaagaagtt tctgccattt atatcggatg      780
gatctccgag cgtaggtata ataccaagga tttcatccac acagctgaag gaccgcgata      840
aagcggactt attcaatagc gtcaccagcg tccgcgtcgg tatcaggaca acagaactgg      900
tacatagaaa tgagctcggt tatcaagccc gcccgcgtcat ccttggtaac tagctgggac      960
gccacatcaa ttggcgacga cgagccttgg agcagacgac ggtcttcgcc gcgtagttta     1020
tcctgtagat c                                     1031
```

<210> 38

<211> 17

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 56

<400> 38

Val Thr Lys Asp Asp Gly Ala Gly Leu Ile Asn Glu Leu Ile Ser Met
1 5 10 15

Tyr

<210> 39

<211> 298

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 56

<220>

<221> misc_feature

<222> (147)..(147)

<223> unknown amino acid

<400> 39

Ser Ser Val Val Leu Ile Pro Thr Arg Thr Leu Val Thr Leu Leu Asn
1 5 10 15

Lys Ser Ala Leu Ser Arg Ser Phe Ser Cys Val Asp Glu Ile Leu Gly
20 25 30

Ile Ile Pro Thr Leu Gly Asp Pro Ser Asp Ile Asn Gly Arg Asn Phe
35 40 45

Phe His His His Val Ile Ala Leu Gly Lys Lys Arg Thr Lys Ser Leu
50 55 60

Glu Glu Arg Asp Asn Ile Asn Ser Leu Leu Ser Asp Ser Leu Asp Leu
65 70 75 80

Glu Ala Ala Ile Pro Pro Glu Pro Asn Thr Arg Leu Val Gly Ala Phe
85 90 95

Gly Pro Asp Gly Val Asn Ser Asp Asp Ser Pro Ala Pro Leu Ser His
100 105 110

Ile Leu Gln Gln Leu Pro Ala His Leu Arg Pro Ser Leu Leu Gln Arg
115 120 125

Asp Asn Tyr Lys Arg Thr Pro Leu His Tyr Ser Ala Gln Tyr Gly Leu
130 135 140

Cys Glu Xaa Thr Arg Ile Ile Leu Gln Ala Leu Ser Glu Trp Asp Ala
145 150 155 160

Trp Asn Ala Asp Val Ala Ile Asp Asp Ile Asp Val Trp Gly Asp Ser
165 170 175

Glu Asn Leu Thr Pro Leu His Leu Ala Val Ile Gly Thr His Pro Leu
180 185 190

Thr Val Ser Thr Leu Leu Ser Phe Met Asn Pro Glu Lys Ser Leu Asn
195 200 205

Ser Pro Arg Leu Leu His Leu Ala Thr Arg Leu Asn Ser Pro Ser Leu
210 215 220

Leu Asn Ser Leu Leu Ser Ala Lys Gly Phe Asp Ile Asp Tyr Gln Glu
225 230 235 240

Pro Glu Asn Leu Glu Thr Ala Leu Tyr Val Ala Cys Lys Leu Asp Ile
245 250 255

Tyr Glu Ala Ala Glu Tyr Leu Val Lys Gln Gly Ala Asn Met Glu Leu
260 265 270

Gly Glu Lys Leu Phe Gly Trp Thr Pro Ile Phe Ala Ala Ala Thr Glu
275 280 285

Gly Tyr Ala Arg Ile Val Gln Leu Leu Val
290 295

<210> 40

<211> 4925

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (426) .. (4388)

<223>

<220>

<221> misc_feature

<223> Oligo 56

```

<400> 40
ctcaagggtt ggctgggtct ggaactacgc cagctcatga cgctatccgt gcgtatcatc 60
aatgacctcg ttaaaacgga tcagctatgt cctagttact gaaggccaaa aagaaatata 120
taaaagtaac gctgggtgcct ttccacctta tccacgtgcg gaagctcggg cgttaagggt 180
acgagctacg cttgcagatt gtgtttgagg aaggccaggg tgctcgtgcc tagcattcat 240
tccgcgcaca tgtgtacata tatactagct ttataggcca cacaacaaac agatatgcgc 300
acggctggag atatctcggc atacaccgcg tggcacagtg ttttcccttg acttctgcat 360
attagagggt tgtatagggg atactagctt cagctgggga gcacacagac taggctagga 420
caagt atg aag ttc ggc aag aca ttt ccc aac cat cag gtg ccg gaa tgg 470
    Met Lys Phe Gly Lys Thr Phe Pro Asn His Gln Val Pro Glu Trp
      1           5           10          15

gca cac aag tac gtg aat tac aag ggt ctg aag aag cag atc aag gaa 518
Ala His Lys Tyr Val Asn Tyr Lys Gly Leu Lys Lys Gln Ile Lys Glu
      20           25           30

atc acg ttg gtg cag gat gcg ctg ttc cgc cag gag caa ggc gca gct 566
Ile Thr Leu Val Gln Asp Ala Leu Phe Arg Gln Glu Gln Gly Ala Ala
      35           40           45

tcg cag gac gga ccc gct cgg cgg cgg gga aga gag agc aag gag cag 614
Ser Gln Asp Gly Pro Ala Arg Arg Arg Gly Arg Glu Ser Lys Glu Gln
      50           55           60

tat ctt ggc cat cca gag gtg aag aag ctg ctt gca gca ttt ttc ttt 662
Tyr Leu Gly His Pro Glu Val Lys Lys Leu Leu Ala Ala Phe Phe Phe
      65           70           75

gcc ctg gac cgg gat atc gag aag gtg gac ggt ttc tac aac atg cag 710
Ala Leu Asp Arg Asp Ile Glu Lys Val Asp Gly Phe Tyr Asn Met Gln
      80           85           90           95

ttt atg gag tat gac cgg cgg ctg agg aag ctt cta tca agc gcg cag 758

```

Phe	Met	Glu	Tyr	Asp	Arg	Arg	Leu	Arg	Lys	Leu	Leu	Ser	Ser	Ala	Gln		
				100					105					110			
ctg	gca	gac	atc	acg	tcg	gta	cag	cgc	ggc	gct	acc	ggc	tac	ctg	cac	806	
Leu	Ala	Asp	Ile	Thr	Ser	Val	Gln	Arg	Gly	Ala	Thr	Gly	Tyr	Leu	His		
			115					120					125				
gcg	cca	ctt	ccg	cag	tac	ata	gca	tac	ggg	gag	cgc	gaa	cgg	gat	gga	854	
Ala	Pro	Leu	Pro	Gln	Tyr	Ile	Ala	Tyr	Gly	Glu	Arg	Glu	Arg	Asp	Gly		
		130						135					140				
ttg	cca	gag	cgc	tat	gta	ccg	ccg	cac	gcc	act	gac	atg	tcg	gag	gac	902	
Leu	Pro	Glu	Arg	Tyr	Val	Pro	Pro	His	Ala	Thr	Asp	Met	Ser	Glu	Asp		
	145					150					155						
cta	gcg	gag	gtg	ctg	acg	att	ctg	ctg	gag	ctg	cgg	tcg	cac	ttc	cgc	950	
Leu	Ala	Glu	Val	Leu	Thr	Ile	Leu	Leu	Glu	Leu	Arg	Ser	His	Phe	Arg		
160					165				170						175		
aac	ctg	aag	tgg	tac	ggg	gag	ctc	aac	aag	cgg	gca	ttc	acg	aaa	atc	998	
Asn	Leu	Lys	Trp	Tyr	Gly	Glu	Leu	Asn	Lys	Arg	Ala	Phe	Thr	Lys	Ile		
				180					185					190			
atg	aag	aaa	ctg	gat	aag	aag	gtt	ggc	aca	aac	cag	cag	cac	tcc	tac	1046	
Met	Lys	Lys	Leu	Asp	Lys	Lys	Val	Gly	Thr	Asn	Gln	Gln	His	Ser	Tyr		
			195					200					205				
ttc	cag	gcc	cgc	att	aag	ccc	tta	gaa	ttt	gct	gac	gat	aca	ccg	atc	1094	
Phe	Gln	Ala	Arg	Ile	Lys	Pro	Leu	Glu	Phe	Ala	Asp	Asp	Thr	Pro	Ile		
		210					215					220					
gtc	aag	gcg	cta	gct	acc	att	aat	gag	atc	cta	gat	cgc	atc	tcg	ccc	1142	
Val	Lys	Ala	Leu	Ala	Thr	Ile	Asn	Glu	Ile	Leu	Asp	Arg	Ile	Ser	Pro		
	225					230					235						
tgc	gtg	aag	gat	cta	cag	gat	aaa	cta	cgc	ggc	gaa	gac	cgt	cgt	ctg	1190	
Cys	Val	Lys	Asp	Leu	Gln	Asp	Lys	Leu	Arg	Gly	Glu	Asp	Arg	Arg	Leu		
240					245				250						255		
ctc	caa	ggc	tcg	tcg	tcg	cca	att	gat	gtg	gcg	tcc	cag	cta	gtt	acc	1238	
Leu	Gln	Gly	Ser	Ser	Ser	Pro	Ile	Asp	Val	Ala	Ser	Gln	Leu	Val	Thr		
			260						265					270			
aag	gat	gac	ggg	gcg	ggc	ttg	ata	aac	gag	ctc	att	tct	atg	tac	cgt	1286	
Lys	Asp	Asp	Gly	Ala	Gly	Leu	Ile	Asn	Glu	Leu	Ile	Ser	Met	Tyr	Arg		
			275					280					285				
tct	gtt	gtc	ctg	ata	ccg	acg	cgg	acg	ctg	gtg	acg	cta	ttg	aat	aag	1334	
Ser	Val	Val	Leu	Ile	Pro	Thr	Arg	Thr	Leu	Val	Thr	Leu	Leu	Asn	Lys		
		290					295					300					
tcc	gct	tta	tcg	cgg	tcc	ttc	agc	tgt	gtg	gat	gaa	atc	ctt	ggg	att	1382	
Ser	Ala	Leu	Ser	Arg	Ser	Phe	Ser	Cys	Val	Asp	Glu	Ile	Leu	Gly	Ile		
	305					310					315						
ata	cct	acg	ctc	gga	gat	cca	tcc	gat	ata	aat	ggc	aga	aac	ttc	ttc	1430	
Ile	Pro	Thr	Leu	Gly	Asp	Pro	Ser	Asp	Ile	Asn	Gly	Arg	Asn	Phe	Phe		
320					325				330						335		
cat	cat	cat	gta	att	gca	ctg	ggc	aaa	aag	cgg	acc	aag	agt	ctc	gaa	1478	
His	His	His	Val	Ile	Ala	Leu	Gly	Lys	Lys	Arg	Thr	Lys	Ser	Leu	Glu		
			340					345						350			

gag cgt gat aac ata aac tcg cta ctg agc gac tct ctc gat cta gag Glu Arg Asp Asn Ile Asn Ser Leu Leu Ser Asp Ser Leu Asp Leu Glu 355 360 365	1526
gca gcg ata ccc ccg gaa ccc aat act aga ctt gtt ggt gct ttt ggt Ala Ala Ile Pro Pro Glu Pro Asn Thr Arg Leu Val Gly Ala Phe Gly 370 375 380	1574
ccg gat ggc gtt aac tct gac gac tct cca gct ccc ctg tca cac ata Pro Asp Gly Val Asn Ser Asp Asp Ser Pro Ala Pro Leu Ser His Ile 385 390 395	1622
tta caa cag ttg cca gcc cat tta aga ccg tcg ctg tta caa aga gat Leu Gln Gln Leu Pro Ala His Leu Arg Pro Ser Leu Leu Gln Arg Asp 400 405 410 415	1670
aac tac aag cga acg ccg cta cat tac tct gcc caa tat ggt cta tgt Asn Tyr Lys Arg Thr Pro Leu His Tyr Ser Ala Gln Tyr Gly Leu Cys 420 425 430	1718
gag gtt acc agg ata att ctg cag gct ctg agt gag tgg gat gct tgg Glu Val Thr Arg Ile Ile Leu Gln Ala Leu Ser Glu Trp Asp Ala Trp 435 440 445	1766
aat gca gac gtt gca atc gat gat att gac gtg tgg gga gat tcg gag Asn Ala Asp Val Ala Ile Asp Asp Ile Asp Val Trp Gly Asp Ser Glu 450 455 460	1814
aac ctc aca cca ttg cat ttg gct gtt atc ggg aca cac cct ctt act Asn Leu Thr Pro Leu His Leu Ala Val Ile Gly Thr His Pro Leu Thr 465 470 475	1862
gtc agc acg cta ttg tcg ttt atg aac cca gag aaa tct ctt aac agc Val Ser Thr Leu Leu Ser Phe Met Asn Pro Glu Lys Ser Leu Asn Ser 480 485 490 495	1910
cct cgt ttg cta cac ctc gcg aca aga cta aac tcg ccg tct ttg tta Pro Arg Leu Leu His Leu Ala Thr Arg Leu Asn Ser Pro Ser Leu Leu 500 505 510	1958
aat tct tta ttg tct gct aag ggc ttt gat att gac tat cag gag ccg Asn Ser Leu Leu Ser Ala Lys Gly Phe Asp Ile Asp Tyr Gln Glu Pro 515 520 525	2006
gaa aat ctc gag act gct ctt tat gtt gct tgc aag tta gac atc tat Glu Asn Leu Glu Thr Ala Leu Tyr Val Ala Cys Lys Leu Asp Ile Tyr 530 535 540	2054
gag gcg gcc gaa tat cta gtt aaa cag ggt gca aat atg gaa ctt ggt Glu Ala Ala Glu Tyr Leu Val Lys Gln Gly Ala Asn Met Glu Leu Gly 545 550 555	2102
gag aag ctt ttt ggc tgg acc cca atc ttt gcg gca gcg act gaa ggg Glu Lys Leu Phe Gly Trp Thr Pro Ile Phe Ala Ala Ala Thr Glu Gly 560 565 570 575	2150
tac gcg cgt atc gtc caa tta tta gtg gat cat ggg gcc aaa tat gat Tyr Ala Arg Ile Val Gln Leu Leu Val Asp His Gly Ala Lys Tyr Asp 580 585 590	2198
ctt ttt gat gag agc ggt tgg aca cca atg gag cac gcg gct ctg ccg	2246

Leu Phe Asp Glu Ser Gly Trp Thr Pro Met Glu His Ala Ala Leu Arg	
595 600 605	
gga cat ttg gat atc tct caa ctc att cgc ata aca gat aat aaa gcg	2294
Gly His Leu Asp Ile Ser Gln Leu Ile Arg Ile Thr Asp Asn Lys Ala	
610 615 620	
atc act cgg ccg aag ttc gcc acc gac tgg aac aag tct acc aga cca	2342
Ile Thr Arg Pro Lys Phe Ala Thr Asp Trp Asn Lys Ser Thr Arg Pro	
625 630 635	
aca gaa acc acg aat ggt ttg tta tcc gca tta aca cca tca gaa tct	2390
Thr Glu Thr Thr Asn Gly Leu Leu Ser Ala Leu Thr Pro Ser Glu Ser	
640 645 650 655	
ggg tct aca acc acg ggc tcc gaa aac aag agc tct tct ttg aca cct	2438
Gly Ser Thr Thr Thr Gly Ser Glu Asn Lys Ser Ser Ser Leu Thr Pro	
660 665 670	
agc acc agc aat gaa atg tat gct ctg cca gca cgc tct tca acg tca	2486
Ser Thr Ser Asn Glu Met Tyr Ala Leu Pro Ala Arg Ser Ser Thr Ser	
675 680 685	
ata gac aag ata tct gaa cca aac aaa gga aac cac agg aag gtt ttg	2534
Ile Asp Lys Ile Ser Glu Pro Asn Lys Gly Asn His Arg Lys Val Leu	
690 695 700	
aag tcc cag cta agc cac ggt aag gta caa act att aag gac aca caa	2582
Lys Ser Gln Leu Ser His Gly Lys Val Gln Thr Ile Lys Asp Thr Gln	
705 710 715	
ttg ccg cag cag ccg att aag tct ttc ggt cat agc ttc ctg cag aaa	2630
Leu Pro Gln Gln Pro Ile Lys Ser Phe Gly His Ser Phe Leu Gln Lys	
720 725 730 735	
gac gag tct gtt atc ctt ttg act ttg ggc act aat gat aat cgc agt	2678
Asp Glu Ser Val Ile Leu Leu Thr Leu Gly Thr Asn Asp Asn Arg Ser	
740 745 750	
acc ata cct gct gtt tct ctg aat aag gtt cct gtt gct aag gct tcc	2726
Thr Ile Pro Ala Val Ser Leu Asn Lys Val Pro Val Ala Lys Ala Ser	
755 760 765	
tca acg gag cta gat aca gcg tta tct ttg ttg gtt aca tgc atg gat	2774
Ser Thr Glu Leu Asp Thr Ala Leu Ser Leu Leu Val Thr Cys Met Asp	
770 775 780	
aat ttg gat gca gaa cct gtt atg ctg gac ctt cca ttg cat gag aac	2822
Asn Leu Asp Ala Glu Pro Val Met Leu Asp Leu Pro Leu His Glu Asn	
785 790 795	
ttg gat tcg gtc act ttt aaa gtc cca tac aag aaa gat tcc tct tat	2870
Leu Asp Ser Val Thr Phe Lys Val Pro Tyr Lys Lys Asp Ser Ser Tyr	
800 805 810 815	
act ata ttt ttt gat atc gtt ccc acc tat ggc tat tca atg gca aac	2918
Thr Ile Phe Phe Asp Ile Val Pro Thr Tyr Gly Tyr Ser Met Ala Asn	
820 825 830	
atg aac cgc gaa aat tcc tct ggt atg cat tcg aat gtt ggt aat agt	2966
Met Asn Arg Glu Asn Ser Ser Gly Met His Ser Asn Val Gly Asn Ser	
835 840 845	

act ggc ccc gcg tac cta gac gca caa gtg ggc cag tgc ggc tct cgt	3014
Thr Gly Pro Ala Tyr Leu Asp Ala Gln Val Gly Gln Cys Gly Ser Arg	
850 855 860	
tta cac tat gac cag ctt ggg aga gat aca cca aac act tac gat cag	3062
Leu His Tyr Asp Gln Leu Gly Arg Asp Thr Pro Asn Thr Tyr Asp Gln	
865 870 875	
cgt tcg cgg cac caa gca tcc caa caa aag gaa caa att gcc acg aag	3110
Arg Ser Arg His Gln Ala Ser Gln Gln Lys Glu Gln Ile Ala Thr Lys	
880 885 890 895	
aag caa agt aag ata tta ggg agg gca gtt gct tta ttg gac tcc gcg	3158
Lys Gln Ser Lys Ile Leu Gly Arg Ala Val Ala Leu Leu Asp Ser Ala	
900 905 910	
cca act tcg gtg ggc ccc aac agg cgc tct att gcg gag gct atc act	3206
Pro Thr Ser Val Gly Pro Asn Arg Arg Ser Ile Ala Glu Ala Ile Thr	
915 920 925	
ata cct atc att ggg agc gac aca ctt gaa gtc ctc ggg att att cga	3254
Ile Pro Ile Ile Gly Ser Asp Thr Leu Glu Val Leu Gly Ile Ile Arg	
930 935 940	
ttt gac ttc ctc gta gtt act ccg ttt gtc cac aag aat tta tca gtt	3302
Phe Asp Phe Leu Val Val Thr Pro Phe Val His Lys Asn Leu Ser Val	
945 950 955	
gga cct gct gaa acg tat tgg aaa tcg ctg gtt tcg acc cgg gtg atc	3350
Gly Pro Ala Glu Thr Tyr Trp Lys Ser Leu Val Ser Thr Arg Val Ile	
960 965 970 975	
gga cac aga ggc ttg ggc aaa aac atg aac acg aac aag tcg ttg cag	3398
Gly His Arg Gly Leu Gly Lys Asn Met Asn Thr Asn Lys Ser Leu Gln	
980 985 990	
ctt ggc gag aat acg gtg gag tcc ttt atc gca gct gcc tca ttg ggt	3446
Leu Gly Glu Asn Thr Val Glu Ser Phe Ile Ala Ala Ala Ser Leu Gly	
995 1000 1005	
gcc tcg tat gtc gaa ttc gat gtc cag ttg aca aag gat aac att	3491
Ala Ser Tyr Val Glu Phe Asp Val Gln Leu Thr Lys Asp Asn Ile	
1010 1015 1020	
ccc gtt gtg tac cac gat ttt ctt gtt gca gaa tcg ggt gtc gat	3536
Pro Val Val Tyr His Asp Phe Leu Val Ala Glu Ser Gly Val Asp	
1025 1030 1035	
att ccc atg cat gag ctc act ttg gaa cag ttt ctc gat ctg aac	3581
Ile Pro Met His Glu Leu Thr Leu Glu Gln Phe Leu Asp Leu Asn	
1040 1045 1050	
ggg gag cgt caa agg cat cag gac gcg aga gag gcc cac aga aac	3626
Gly Glu Arg Gln Arg His Gln Asp Ala Arg Glu Ala His Arg Asn	
1055 1060 1065	
cac agg agc ccg aac ggc aga cgc ttg tcc atg gat gac agc tct	3671
His Arg Ser Pro Asn Gly Arg Arg Leu Ser Met Asp Asp Ser Ser	
1070 1075 1080	
gct gag cta att aag cgg tcc ctg atg atg cgg ggc gat gaa gac	3716

Ala	Glu	Leu	Ile	Lys	Arg	Ser	Leu	Met	Met	Arg	Gly	Asp	Glu	Asp		
		1085					1090					1095				
cgc	aca	gct	aga	gac	ctt	aac	aca	ata	tac	ggc	gac	cgt	atg	cgg	3761	
Arg	Thr	Ala	Arg	Asp	Leu	Asn	Thr	Ile	Tyr	Gly	Asp	Arg	Met	Arg		
		1100					1105					1110				
ctg	acc	aga	acg	ttc	aag	aag	aat	gcc	ttc	aaa	gcc	aac	tcc	agg	3806	
Leu	Thr	Arg	Thr	Phe	Lys	Lys	Asn	Ala	Phe	Lys	Ala	Asn	Ser	Arg		
		1115					1120					1125				
ggc	cat	gcc	att	gca	tct	agt	ttt	gtt	acg	cta	aaa	gag	ctg	ttc	3851	
Gly	His	Ala	Ile	Ala	Ser	Ser	Phe	Val	Thr	Leu	Lys	Glu	Leu	Phe		
		1130					1135					1140				
aag	aag	atc	ccc	cag	aat	gtt	ggc	ttc	aac	atc	gag	tgc	aag	tat	3896	
Lys	Lys	Ile	Pro	Gln	Asn	Val	Gly	Phe	Asn	Ile	Glu	Cys	Lys	Tyr		
		1145					1150					1155				
cca	atg	gta	gac	gag	gcc	gag	gag	gaa	gac	atc	ggc	ccg	atc	gcc	3941	
Pro	Met	Val	Asp	Glu	Ala	Glu	Glu	Glu	Asp	Ile	Gly	Pro	Ile	Ala		
		1160					1165					1170				
gtg	gaa	atg	aac	cat	tgg	atc	gat	acc	gtg	ctg	gag	gtt	gtc	tac	3986	
Val	Glu	Met	Asn	His	Trp	Ile	Asp	Thr	Val	Leu	Glu	Val	Val	Tyr		
		1175					1180					1185				
gac	aac	gtc	gag	ggc	cgt	gac	gtc	atc	ttt	tcg	tcg	ttt	cag	cca	4031	
Asp	Asn	Val	Glu	Gly	Arg	Asp	Val	Ile	Phe	Ser	Ser	Phe	Gln	Pro		
		1190					1195					1200				
gac	gtg	tgc	ctc	atg	ctc	tcc	cta	aag	cag	ccc	tcc	ttc	ccg	atc	4076	
Asp	Val	Cys	Leu	Met	Leu	Ser	Leu	Lys	Gln	Pro	Ser	Phe	Pro	Ile		
		1205					1210					1215				
ctg	ttc	cta	acg	gaa	ggc	ggg	acc	gcg	aag	cgc	tgc	gac	atc	cgc	4121	
Leu	Phe	Leu	Thr	Glu	Gly	Gly	Thr	Ala	Lys	Arg	Cys	Asp	Ile	Arg		
		1220					1225					1230				
gcg	gcg	tcg	ctg	cag	aat	gcc	atc	cgc	ttc	gcg	cac	cgc	tgg	aac	4166	
Ala	Ala	Ser	Leu	Gln	Asn	Ala	Ile	Arg	Phe	Ala	His	Arg	Trp	Asn		
		1235					1240					1245				
ctg	ctg	ggc	atc	gtc	tcg	gcc	gcg	gcg	cca	atc	gtc	atc	gcg	ccc	4211	
Leu	Leu	Gly	Ile	Val	Ser	Ala	Ala	Ala	Pro	Ile	Val	Ile	Ala	Pro		
		1250					1255					1260				
cgc	ctg	gcc	cag	atc	gtc	aag	tcc	agt	ggc	ctt	gtg	tgc	gtg	acg	4256	
Arg	Leu	Ala	Gln	Ile	Val	Lys	Ser	Ser	Gly	Leu	Val	Cys	Val	Thr		
		1265					1270					1275				
tac	ggc	gtc	gag	aac	aac	gac	ccc	gag	atc	gcc	cgc	gtc	gag	atg	4301	
Tyr	Gly	Val	Glu	Asn	Asn	Asp	Pro	Glu	Ile	Ala	Arg	Val	Glu	Met		
		1280					1285					1290				
gac	gcc	ggc	gtc	gac	gcg	gtc	atc	gtc	gac	agc	gtg	ctc	gcg	gtc	4346	
Asp	Ala	Gly	Val	Asp	Ala	Val	Ile	Val	Asp	Ser	Val	Leu	Ala	Val		
		1295					1300					1305				
cgc	aag	ggc	ctc	acc	cgc	gag	gca	cag	gac	gcc	gac	acg	ctc		4388	
Arg	Lys	Gly	Leu	Thr	Arg	Glu	Ala	Gln	Asp	Ala	Asp	Thr	Leu			
		1310					1315					1320				

taagaacaaa tcgtgattca ctaagcagaa gatatatcgt tatgtagagc ctgccgggtc 4448
ccaggctaatt ccacccattg ccgtcacaaa atcgtaaccg agatgcgccg cccgtgcccc 4508
cgcgcgccgc cctgcagact cgcgcccgcg ctctgcatcg tcagcttctc caagctgact 4568
aacttgctcg tcacgtgatc atccttgaac atttccccga tctgggtgctt gcccgcgctca 4628
tccgtgaaca cctgcaacgt cttcgtctgc gccgcctcgt tgatgtggtg gtactcgtcc 4688
gggaagggct ggatgtgctc gaccgtcagt tcgttcaatt ggatgaactt catcactgtg 4748
aggttcgaca cctcgactgg aatattcgaa actatcacac tgtgcgccat ggggactctg 4808
tttgactgc cgcccttggt gcgccactga tgctgccacc gccacgtcca gcttatatac 4868
ccgctgggaa agttgtcgaa ttgtgaaata gccgacgcta tcggtaccat ttacacc 4925

<210> 41

<211> 1321

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 56

<400> 41

Met Lys Phe Gly Lys Thr Phe Pro Asn His Gln Val Pro Glu Trp Ala
1 5 10 15

His Lys Tyr Val Asn Tyr Lys Gly Leu Lys Lys Gln Ile Lys Glu Ile
20 25 30

Thr Leu Val Gln Asp Ala Leu Phe Arg Gln Glu Gln Gly Ala Ala Ser
35 40 45

Gln Asp Gly Pro Ala Arg Arg Arg Gly Arg Glu Ser Lys Glu Gln Tyr
50 55 60

Leu Gly His Pro Glu Val Lys Lys Leu Leu Ala Ala Phe Phe Phe Ala
65 70 75 80

Leu Asp Arg Asp Ile Glu Lys Val Asp Gly Phe Tyr Asn Met Gln Phe
85 90 95

Met Glu Tyr Asp Arg Arg Leu Arg Lys Leu Leu Ser Ser Ala Gln Leu
 100 105 110

Ala Asp Ile Thr Ser Val Gln Arg Gly Ala Thr Gly Tyr Leu His Ala
 115 120 125

Pro Leu Pro Gln Tyr Ile Ala Tyr Gly Glu Arg Glu Arg Asp Gly Leu
 130 135 140

Pro Glu Arg Tyr Val Pro Pro His Ala Thr Asp Met Ser Glu Asp Leu
 145 150 155 160

Ala Glu Val Leu Thr Ile Leu Leu Glu Leu Arg Ser His Phe Arg Asn
 165 170 175

Leu Lys Trp Tyr Gly Glu Leu Asn Lys Arg Ala Phe Thr Lys Ile Met
 180 185 190

Lys Lys Leu Asp Lys Lys Val Gly Thr Asn Gln Gln His Ser Tyr Phe
 195 200 205

Gln Ala Arg Ile Lys Pro Leu Glu Phe Ala Asp Asp Thr Pro Ile Val
 210 215 220

Lys Ala Leu Ala Thr Ile Asn Glu Ile Leu Asp Arg Ile Ser Pro Cys
 225 230 235 240

Val Lys Asp Leu Gln Asp Lys Leu Arg Gly Glu Asp Arg Arg Leu Leu
 245 250 255

Gln Gly Ser Ser Ser Pro Ile Asp Val Ala Ser Gln Leu Val Thr Lys
 260 265 270

Asp Asp Gly Ala Gly Leu Ile Asn Glu Leu Ile Ser Met Tyr Arg Ser
 275 280 285

Val Val Leu Ile Pro Thr Arg Thr Leu Val Thr Leu Leu Asn Lys Ser
 290 295 300

Ala Leu Ser Arg Ser Phe Ser Cys Val Asp Glu Ile Leu Gly Ile Ile
 305 310 315 320

Pro Thr Leu Gly Asp Pro Ser Asp Ile Asn Gly Arg Asn Phe Phe His
 325 330 335

His His Val Ile Ala Leu Gly Lys Lys Arg Thr Lys Ser Leu Glu Glu

340	345	350
Arg Asp Asn Ile Asn Ser Leu Leu Ser Asp Ser Leu Asp Leu Glu Ala		
355	360	365
Ala Ile Pro Pro Glu Pro Asn Thr Arg Leu Val Gly Ala Phe Gly Pro		
370	375	380
Asp Gly Val Asn Ser Asp Asp Ser Pro Ala Pro Leu Ser His Ile Leu		
385	390	400
Gln Gln Leu Pro Ala His Leu Arg Pro Ser Leu Leu Gln Arg Asp Asn		
405	410	415
Tyr Lys Arg Thr Pro Leu His Tyr Ser Ala Gln Tyr Gly Leu Cys Glu		
420	425	430
Val Thr Arg Ile Ile Leu Gln Ala Leu Ser Glu Trp Asp Ala Trp Asn		
435	440	445
Ala Asp Val Ala Ile Asp Asp Ile Asp Val Trp Gly Asp Ser Glu Asn		
450	455	460
Leu Thr Pro Leu His Leu Ala Val Ile Gly Thr His Pro Leu Thr Val		
465	470	475
Ser Thr Leu Leu Ser Phe Met Asn Pro Glu Lys Ser Leu Asn Ser Pro		
485	490	495
Arg Leu Leu His Leu Ala Thr Arg Leu Asn Ser Pro Ser Leu Leu Asn		
500	505	510
Ser Leu Leu Ser Ala Lys Gly Phe Asp Ile Asp Tyr Gln Glu Pro Glu		
515	520	525
Asn Leu Glu Thr Ala Leu Tyr Val Ala Cys Lys Leu Asp Ile Tyr Glu		
530	535	540
Ala Ala Glu Tyr Leu Val Lys Gln Gly Ala Asn Met Glu Leu Gly Glu		
545	550	555
Lys Leu Phe Gly Trp Thr Pro Ile Phe Ala Ala Ala Thr Glu Gly Tyr		
565	570	575
Ala Arg Ile Val Gln Leu Leu Val Asp His Gly Ala Lys Tyr Asp Leu		
580	585	590

Phe Asp Glu Ser Gly Trp Thr Pro Met Glu His Ala Ala Leu Arg Gly
 595 600 605

His Leu Asp Ile Ser Gln Leu Ile Arg Ile Thr Asp Asn Lys Ala Ile
 610 615 620

Thr Arg Pro Lys Phe Ala Thr Asp Trp Asn Lys Ser Thr Arg Pro Thr
 625 630 635 640

Glu Thr Thr Asn Gly Leu Leu Ser Ala Leu Thr Pro Ser Glu Ser Gly
 645 650 655

Ser Thr Thr Thr Gly Ser Glu Asn Lys Ser Ser Ser Leu Thr Pro Ser
 660 665 670

Thr Ser Asn Glu Met Tyr Ala Leu Pro Ala Arg Ser Ser Thr Ser Ile
 675 680 685

Asp Lys Ile Ser Glu Pro Asn Lys Gly Asn His Arg Lys Val Leu Lys
 690 695 700

Ser Gln Leu Ser His Gly Lys Val Gln Thr Ile Lys Asp Thr Gln Leu
 705 710 715 720

Pro Gln Gln Pro Ile Lys Ser Phe Gly His Ser Phe Leu Gln Lys Asp
 725 730 735

Glu Ser Val Ile Leu Leu Thr Leu Gly Thr Asn Asp Asn Arg Ser Thr
 740 745 750

Ile Pro Ala Val Ser Leu Asn Lys Val Pro Val Ala Lys Ala Ser Ser
 755 760 765

Thr Glu Leu Asp Thr Ala Leu Ser Leu Leu Val Thr Cys Met Asp Asn
 770 775 780

Leu Asp Ala Glu Pro Val Met Leu Asp Leu Pro Leu His Glu Asn Leu
 785 790 795 800

Asp Ser Val Thr Phe Lys Val Pro Tyr Lys Lys Asp Ser Ser Tyr Thr
 805 810 815

Ile Phe Phe Asp Ile Val Pro Thr Tyr Gly Tyr Ser Met Ala Asn Met
 820 825 830

Asn Arg Glu Asn Ser Ser Gly Met His Ser Asn Val Gly Asn Ser Thr
 835 840 845

Gly Pro Ala Tyr Leu Asp Ala Gln Val Gly Gln Cys Gly Ser Arg Leu
 850 855 860

His Tyr Asp Gln Leu Gly Arg Asp Thr Pro Asn Thr Tyr Asp Gln Arg
 865 870 875 880

Ser Arg His Gln Ala Ser Gln Gln Lys Glu Gln Ile Ala Thr Lys Lys
 885 890 895

Gln Ser Lys Ile Leu Gly Arg Ala Val Ala Leu Leu Asp Ser Ala Pro
 900 905 910

Thr Ser Val Gly Pro Asn Arg Arg Ser Ile Ala Glu Ala Ile Thr Ile
 915 920 925

Pro Ile Ile Gly Ser Asp Thr Leu Glu Val Leu Gly Ile Ile Arg Phe
 930 935 940

Asp Phe Leu Val Val Thr Pro Phe Val His Lys Asn Leu Ser Val Gly
 945 950 955 960

Pro Ala Glu Thr Tyr Trp Lys Ser Leu Val Ser Thr Arg Val Ile Gly
 965 970 975

His Arg Gly Leu Gly Lys Asn Met Asn Thr Asn Lys Ser Leu Gln Leu
 980 985 990

Gly Glu Asn Thr Val Glu Ser Phe Ile Ala Ala Ala Ser Leu Gly Ala
 995 1000 1005

Ser Tyr Val Glu Phe Asp Val Gln Leu Thr Lys Asp Asn Ile Pro
 1010 1015 1020

Val Val Tyr His Asp Phe Leu Val Ala Glu Ser Gly Val Asp Ile
 1025 1030 1035

Pro Met His Glu Leu Thr Leu Glu Gln Phe Leu Asp Leu Asn Gly
 1040 1045 1050

Glu Arg Gln Arg His Gln Asp Ala Arg Glu Ala His Arg Asn His
 1055 1060 1065

Arg Ser Pro Asn Gly Arg Arg Leu Ser Met Asp Asp Ser Ser Ala
 1070 1075 1080

Glu	Leu	Ile	Lys	Arg	Ser	Leu	Met	Met	Arg	Gly	Asp	Glu	Asp	Arg
1085						1090					1095			
Thr	Ala	Arg	Asp	Leu	Asn	Thr	Ile	Tyr	Gly	Asp	Arg	Met	Arg	Leu
1100						1105					1110			
Thr	Arg	Thr	Phe	Lys	Lys	Asn	Ala	Phe	Lys	Ala	Asn	Ser	Arg	Gly
1115						1120					1125			
His	Ala	Ile	Ala	Ser	Ser	Phe	Val	Thr	Leu	Lys	Glu	Leu	Phe	Lys
1130						1135					1140			
Lys	Ile	Pro	Gln	Asn	Val	Gly	Phe	Asn	Ile	Glu	Cys	Lys	Tyr	Pro
1145						1150					1155			
Met	Val	Asp	Glu	Ala	Glu	Glu	Glu	Asp	Ile	Gly	Pro	Ile	Ala	Val
1160						1165					1170			
Glu	Met	Asn	His	Trp	Ile	Asp	Thr	Val	Leu	Glu	Val	Val	Tyr	Asp
1175						1180					1185			
Asn	Val	Glu	Gly	Arg	Asp	Val	Ile	Phe	Ser	Ser	Phe	Gln	Pro	Asp
1190						1195					1200			
Val	Cys	Leu	Met	Leu	Ser	Leu	Lys	Gln	Pro	Ser	Phe	Pro	Ile	Leu
1205						1210					1215			
Phe	Leu	Thr	Glu	Gly	Gly	Thr	Ala	Lys	Arg	Cys	Asp	Ile	Arg	Ala
1220						1225					1230			
Ala	Ser	Leu	Gln	Asn	Ala	Ile	Arg	Phe	Ala	His	Arg	Trp	Asn	Leu
1235						1240					1245			
Leu	Gly	Ile	Val	Ser	Ala	Ala	Ala	Pro	Ile	Val	Ile	Ala	Pro	Arg
1250						1255					1260			
Leu	Ala	Gln	Ile	Val	Lys	Ser	Ser	Gly	Leu	Val	Cys	Val	Thr	Tyr
1265						1270					1275			
Gly	Val	Glu	Asn	Asn	Asp	Pro	Glu	Ile	Ala	Arg	Val	Glu	Met	Asp
1280						1285					1290			
Ala	Gly	Val	Asp	Ala	Val	Ile	Val	Asp	Ser	Val	Leu	Ala	Val	Arg
1295						1300					1305			

Lys Gly Leu Thr Arg Glu Ala Gln Asp Ala Asp Thr Leu
 1310 1315 1320

<210> 42

<211> 1078

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 167

<220>

<221> misc_feature

<222> (436)..(436)

<223> unknown nucleotide

<220>

<221> misc_feature

<222> (584)..(584)

<223> unknown nucleotide

<400> 42

```

gacacctoga gcgtagacga agaaaaggtc aaacgcctga aatcgacctt taccttttag      60
ctctcaacta atccgggtcgt tatcaatcgt gtaattctat accttaaata gactacctaa    120
ccagcacgca gtgtcccata cagggccttgc agtatggtgc ctaggtctca ttcgttgtat    180
tcttttgctc tgcagctttc ctttgggccc atgcagcgta agccctgccg ttcgcaatca    240
acggattatc atgcttgaaa tagtatgggc cagagagcgc tgaatattgt ctcttgtcta    300
agttatacac tctttccatt tcttccttat agacgtcctc gataatttta acatctccct    360
catctatctt cgtgttggtg gaggcggtgt ataaatctgc atcgaatggc cgaatgcaat    420
tgccgacata gaaganggtg gatgtgtgcc cctcacccaa ttcaaactcg gatttgggggt    480
tacgcamccg ccctagatgt tccataagcc gccgtgctag ctgcggatgg tctggctgct    540
gcttttgctg atgtggcctg tcactctctg tttcctctcc acctgagac gccatagatg    600

```

gcccaacaga gtttcgtggc acgtcaatcc ctccgtctca tggctgggta tgcgtcgcag 660
 tccggaatgg ggacagttat atacgcagca aactgcagat cgacccacgc ggctgtgcc 720
 cacgacacac attaccgtag tcgcaggcgc ttatgctcac aaaagaaagc cctgtatata 780
 aggtcctaata tatatagagc tcttacaaca ctttctgttt catgaagaag gtgcgcaggt 840
 gggtgacctg ccagaagcat gtgccgcaa gcacaaacaa cgtcatgacc gcacaccata 900
 caaccacga attggcgctc tcggacgcat ctgcggaatt gggcctccct ctactgcac 960
 caacctgctg ctctgttgct ggatcgccct cacttctcg ccgagctgta gcacgcgcga 1020
 atgcaggtac tgcagcttgt ccttcttctt gggatccagc agtaggtccg atccgatc 1078

<210> 43

<211> 42

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 167

<400> 43

Asp	Ala	Ser	Glu	Ser	Ala	Asn	Ser	Trp	Val	Val	Trp	Cys	Ala	Val	Met
1				5					10					15	

Thr	Leu	Phe	Val	Leu	Gly	Gly	Thr	Cys	Phe	Trp	Gln	Val	Asn	His	Leu
			20					25					30		

Arg	Thr	Phe	Phe	Met	Lys	Gln	Lys	Val	Leu
			35				40		

<210> 44

<211> 2549

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (563) .. (1216)

<223>

<220>

<221> misc_feature

<223> Oligo 167

<400> 44

tatgccccgc ggctgcgggc ccgccagcac ccggcacatc caaacccggcc catcaaacat	60
gtactcctgg ttctgccgcc ccaccgccgc gaggtccac atcttcactc tccaatcccg	120
cgagctgctc cataggtacc gtccacattg cgaccagcc accgactgca ccgctcgtac	180
atggccgccc ccatggtttc ccagcaccga aataggcttc ttcgtatcca tgcgtagat	240
caccagagac ccgttggagc atccaactgc caggtagtca ccgcctgggc tgaactccag	300
acaatcgc at tgcagcggga tttcaaactg atatgttagc gtctcagggt attccttcag	360
tacactgaat ggatcctgca acaaccgatt tgccatcgcc ccttccccgc ctccaagcct	420
ccaacctgat gctttcctcc aggctgtacc tggaaagtaa aaaaaattc gaacagcgaa	480
ccagatatgg acaccgaatt gaagatcttg ttccccctgg aatctcagta tatcgatcga	540
tcaggcgctg aaggcttcca gg atg aac ttg aag aca ttt tta cag gta ctg	592
Met Asn Leu Lys Thr Phe Leu Gln Val Leu	
1 5 10	
gtg ctc ggc atg ttg ccg ttg cag gtc agt ggg ttc tac ttc tat gtc	640
Val Leu Gly Met Leu Pro Leu Gln Val Ser Gly Phe Tyr Phe Tyr Val	
15 20 25	
aat ggt gga gac cgt aag tgc ttc cac agt gaa tta atg aaa gat gcg	688
Asn Gly Gly Asp Arg Lys Cys Phe His Ser Glu Leu Met Lys Asp Ala	
30 35 40	
gtt ttg aac ctg aag tac aac gtg caa tcg tac gac tcc cag tct gga	736
Val Leu Asn Leu Lys Tyr Asn Val Gln Ser Tyr Asp Ser Gln Ser Gly	
45 50 55	
acg tac cgg gac atg cgg gac ggt gaa cta atg atg atg atc gac ata	784
Thr Tyr Arg Asp Met Arg Asp Gly Glu Leu Met Met Met Ile Asp Ile	
60 65 70	
gag gag gtg ttt gac gac aac cac cgt gtt atg cac cag aaa ttg ccg	832
Glu Glu Val Phe Asp Asp Asn His Arg Val Met His Gln Lys Leu Pro	
75 80 85 90	
gcg tac ggt aca act acg ttc aac gcg gtt gac tct ggc gag cac aag	880

Ala	Tyr	Gly	Thr	Thr	Phe	Asn	Ala	Val	Asp	Ser	Gly	Glu	His	Lys			
95				100				105									
gtt	tgc	gtt	cag	ccg	cag	ttc	cag	ggg	tgg	atg	gcg	cgc	acc	aag	aca	928	
Val	Cys	Val	Gln	Pro	Gln	Phe	Gln	Gly	Trp	Met	Ala	Arg	Thr	Lys	Thr		
110				115				120									
aag	atc	acg	ctt	gac	ttt	gag	atc	gga	tcg	gac	cta	ctg	ctg	gat	ccc	976	
Lys	Ile	Thr	Leu	Asp	Phe	Glu	Ile	Gly	Ser	Asp	Leu	Leu	Leu	Asp	Pro		
125				130				135									
aag	aag	aag	gac	aag	ctg	cag	tac	ctg	cat	tcg	cgc	gtg	cta	cag	ctc	1024	
Lys	Lys	Lys	Asp	Lys	Leu	Gln	Tyr	Leu	His	Ser	Arg	Val	Leu	Gln	Leu		
140				145				150									
ggc	gag	aag	gtg	aag	gcg	atc	cgc	aac	gag	cag	cgg	ttg	gtg	cgt	gag	1072	
Gly	Glu	Lys	Val	Lys	Ala	Ile	Arg	Asn	Glu	Gln	Arg	Leu	Val	Arg	Glu		
155				160				165				170					
agg	gag	gcc	caa	ttc	cgc	gat	gcg	tcc	gag	agc	gcc	aat	tcg	tgg	gtt	1120	
Arg	Glu	Ala	Gln	Phe	Arg	Asp	Ala	Ser	Glu	Ser	Ala	Asn	Ser	Trp	Val		
175				180				185									
gta	tgg	tgt	gcg	gtc	atg	acg	ttg	ttt	gtg	ctt	ggc	ggc	aca	tgc	ttc	1168	
Val	Trp	Cys	Ala	Val	Met	Thr	Leu	Phe	Val	Leu	Gly	Gly	Thr	Cys	Phe		
190				195				200									
tgg	cag	gtc	aac	cac	ctg	cgc	acc	ttc	ttc	atg	aaa	cag	aaa	gtg	ttg	1216	
Trp	Gln	Val	Asn	His	Leu	Arg	Thr	Phe	Phe	Met	Lys	Gln	Lys	Val	Leu		
205				210				215									
taagagctct			atataattag			gaccttatat			acagggcttt			cttttgtgag			cataagcgcc		1276
tgcgactacg			gtaatgtgtg			tcgtgggcac			aggccgcgtg			ggtcgatctg			cagtttgctg		1336
cgtatataac			tgtccccatt			ccggactgcg			acgcataacc			agccatgaga			cggagggatt		1396
gacgtgccac			gaaactctgt			tgggccatct			atggcgtctc			aggggtggaga			ggaaacagag		1456
agtgacaggc			cacatcagca			aaagcagcag			ccagaccatc			cgcagctagc			acggcggcctt		1516
atggaacatc			tagggcggtt			gcgtaacccc			aaatccgagt			ttgaattggg			tgagggggcac		1576
acatccaccc			tcttctatgt			cggcaattgc			attcggccat			tcgatgcaga			tttatacacc		1636
gcctccaaca			acacgaagat			agatgagggga			gatgttaaaa			ttatcgagga			cgtctataag		1696
gaagaaatgg			aaagagtgtg			taacttagac			aagagacaat			attcagcgct			ctctggccca		1756
tactatttca			agcatgataa			tccgttgatt			gcgaacggca			gggcttacgc			tgcattgggcc		1816
caaaggaaaag			ctgcagagca			aaagaataca			acgaatgaga			cctagcacca			tacgcaagcc		1876
ctgtatggga			cactgcgtgc			tggttaggtg			gtctatttaa			ggtatagaat			tacacgattg		1936
ataacgaccg			gattagttga			gagctaaaag			gtaaagggtcg			atttcaggcg			tttgaccttt		1996
tcttcgtcta			cgctcgaagg			atcaatccca			gatacaggta			aggttccgaa			gtcaaagtgg		2056
aacggaaccg			cttcggcaga			tgctgcagtg			ttctctgcaa			atgaggagtc			atccttgggt		2116
aagtcaggtt			tcgagaagga			aaatgcaaag			gggggtgcggg			atgtcgtatt			gtgcttatct		2176

```

tcacttttat gaacttggga gactttctga atgccattag aaaaattgaa cgacccattc 2236
gactgcaatg agctccccga tatgttattc tccatgggcc cgtaagagt agctccgtgg 2296
gtattagatt cctctacttt ggccagaggt tgaaacgaaa atgtgctggt tgcacaggt 2356
atagtgtgtg gcttctcaga caaagtatcc acattgtggt ctttgggctt ttggaaagaa 2416
aagttaaatt tagagctctg tggaacaggc tgccgaggcg acttttcagg cgacttttca 2476
acagtcacct caacgcgcca gcgatgaagg cctggaaggt ttaaaccggg acactgggtg 2536
tgtagcatga gga 2549

```

<210> 45

<211> 218

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 167

<400> 45

```

Met Asn Leu Lys Thr Phe Leu Gln Val Leu Val Leu Gly Met Leu Pro
1           5           10           15

```

```

Leu Gln Val Ser Gly Phe Tyr Phe Tyr Val Asn Gly Gly Asp Arg Lys
20           25           30

```

```

Cys Phe His Ser Glu Leu Met Lys Asp Ala Val Leu Asn Leu Lys Tyr
35           40           45

```

```

Asn Val Gln Ser Tyr Asp Ser Gln Ser Gly Thr Tyr Arg Asp Met Arg
50           55           60

```

```

Asp Gly Glu Leu Met Met Met Ile Asp Ile Glu Glu Val Phe Asp Asp
65           70           75           80

```

```

Asn His Arg Val Met His Gln Lys Leu Pro Ala Tyr Gly Thr Thr Thr
85           90           95

```

```

Phe Asn Ala Val Asp Ser Gly Glu His Lys Val Cys Val Gln Pro Gln
100          105          110

```

Phe Gln Gly Trp Met Ala Arg Thr Lys Thr Lys Ile Thr Leu Asp Phe
115 120 125

Glu Ile Gly Ser Asp Leu Leu Leu Asp Pro Lys Lys Lys Asp Lys Leu
130 135 140

Gln Tyr Leu His Ser Arg Val Leu Gln Leu Gly Glu Lys Val Lys Ala
145 150 155 160

Ile Arg Asn Glu Gln Arg Leu Val Arg Glu Arg Glu Ala Gln Phe Arg
165 170 175

Asp Ala Ser Glu Ser Ala Asn Ser Trp Val Val Trp Cys Ala Val Met
180 185 190

Thr Leu Phe Val Leu Gly Gly Thr Cys Phe Trp Gln Val Asn His Leu
195 200 205

Arg Thr Phe Phe Met Lys Gln Lys Val Leu
210 215