

(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
6. März 2003 (06.03.2003)

PCT

(10) Internationale Veröffentlichungsnummer
WO 03/018813 A2

(51) Internationale Patentklassifikation⁷: **C12N 15/52**,
9/00, 15/31, 15/81, 1/19, C12Q 1/25, C12P 25/00

(21) Internationales Aktenzeichen: PCT/EP02/09454

(22) Internationales Anmeldedatum:
23. August 2002 (23.08.2002)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:

101 41 267.3	23. August 2001 (23.08.2001)	DE
101 41 268.1	23. August 2001 (23.08.2001)	DE
101 41 280.0	23. August 2001 (23.08.2001)	DE
101 41 281.9	23. August 2001 (23.08.2001)	DE
101 41 282.7	23. August 2001 (23.08.2001)	DE
101 41 283.5	23. August 2001 (23.08.2001)	DE
101 41 284.3	23. August 2001 (23.08.2001)	DE
101 41 288.6	23. August 2001 (23.08.2001)	DE
101 41 289.4	23. August 2001 (23.08.2001)	DE
101 41 290.8	23. August 2001 (23.08.2001)	DE
101 41 291.6	23. August 2001 (23.08.2001)	DE
102 09 829.8	6. März 2002 (06.03.2002)	DE
102 16 027.9	11. April 2002 (11.04.2002)	DE
102 16 031.7	11. April 2002 (11.04.2002)	DE
102 16 032.5	11. April 2002 (11.04.2002)	DE
102 21 912.5	16. Mai 2002 (16.05.2002)	DE
102 21 914.1	16. Mai 2002 (16.05.2002)	DE
102 21 917.6	16. Mai 2002 (16.05.2002)	DE
102 21 920.6	16. Mai 2002 (16.05.2002)	DE
102 21 927.3	16. Mai 2002 (16.05.2002)	DE
102 27 797.4	21. Juni 2002 (21.06.2002)	DE
102 27 798.2	21. Juni 2002 (21.06.2002)	DE

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von
US): **BASF AKTIENGESellschaft** [DE/DE]; 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, ES-37001
Salamanca (ES).

(74) Anwälte: **KINZEBACH, Werner** usw.; 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, VC, 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 METABOLISM-ASSOCIATED GENE PRODUCTS OF ASHBYA GOSSYPHII

(54) Bezeichnung: NEUE STOFFWECHSEL-ASSOZIIERTE GENPRODUKTE AUS ASHBYA GOSSYPHII

(57) Abstract: The invention relates to novel polynucleotides of *Ashbya gossypii*; oligonucleotides hybridizing therewith; expression cassettes and vectors containing said polynucleotides; microorganisms transformed therewith; polypeptides coded by said polynucleotides; and the use of the novel polypeptides and polynucleotides as targets in the modulation of metabolism reactions and especially in order to improve vitamin B2 production in microorganisms of the genus *Ashbya*.

(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 Modulation der Stoffwechselreaktionen und insbesondere zur Verbesserung der Vitamin B2-Produktion in Mikroorganismen der Gattung *Ashbya*.



WO 03/018813 A2

Beschreibung

Neue Stoffwechsel-assoziierte Genprodukte aus *Ashbya gossypii*.

- 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 Stoffwechsels und insbesondere der Verbesserung der Vitamin B2-Produktion in Mikroorganismen der Gattung *Ashbya*.

- Vitamin B2 (Riboflavin, Lactoflavin) ist ein alkali- und lichtempfindliches, in Lösung gelbgrün fluoreszierendes Vitamin. Vitamin B2-Mangel kann zu Ektodermschäden, 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 ausgeschieden wird.
- 30

- 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-
- 35

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.

Der Gesamtprozeß, durch den lebende Systeme die zur Ausübung ihrer verschiedenen Funktionen benötigte freie Enthalpie erlangen und verwerten, wird als Stoffwechsel bezeichnet. Die Wege des Stoffwechsels bestehen aus Abfolgen von enzymatischen Reaktionen, die spezifische Produkte liefern. Die Stoffwechselreaktionswege werden häufig in zwei Kategorien unterteilt. Zum einen in die am Abbau beteiligten Wege (Katabolismus) und zum anderen in die an der Biosynthese involvierten Wege (Anabolismus). Bei den katabolischen Wegen werden komplexe Metaboliten exergonisch in einfachere Produkte zerlegt. Die bei diesen Prozessen freigesetzte Freie Enthalpie wird durch die Synthese von in der Zelle universell einsetzbaren, energiereichen Verbindungen (ATP, NADPH) gespeichert. Bei ATP und NADPH handelt es sich um die wichtigsten Freie-Enthalpie-Quellen für viele Biosynthesereaktionen in den anabolischen Stoffwechselwegen. Beispielsweise werden NADPH und ATP bei zahlreichen Syntheseschritten zur Produktion von Feinchemikalien wie im Falle der Riboflavinsynthese benötigt.

In den katabolischen Stoffwechselwegen wird eine große Zahl unterschiedlicher Substanzen (Kohlenhydrate, Lipide und Proteine) in die gleichen Zwischenprodukte überführt. In der Biosynthese findet der umgekehrte Prozeß statt. Relativ wenige Metaboliten dienen als Ausgangssubstanzen für eine große Zahl unterschiedlicher Biosynthesprodukte. Durch den geregelten Eingriff bzw. die Blockierung von Abbauwegen bereits synthetisierter Feinchemikalien kann deren Produktion zusätzlich gesteigert werden.

Einige Produkte und Nebenprodukte von diesen natürlich vorkommenden metabolischen Prozessen finden industrielle Anwendungen in einem breiten Gebiet, wie z.B. in der Nahrungs-, Futtermittel-, kosmetischen und pharmazeutischen Industrie. Zu diese Molekülen, zusammenfassend "Feinchemikalien" genannt, gehören organische Säuren, proteinogene und nicht-

proteinogene Aminosäuren, Nukleotide und Nukleoside, Lipide und Fettsäuren, Diole, Kohlehydrate, aromatische Verbindungen, Vitamine und Cofaktoren, und Enzyme. Die Produktion dieser Substanzen wird meist in großvolumigen Fermentern durchgeführt, in denen die jeweils gewünschten Moleküle in großen Konzentrationen ins Medium ausgeschieden werden. Ein besonders nützlicher Organismus für dieses Verfahren ist der filamentöse Ascomycet *Ashbya gossypii*.

Eine Veränderung von Proteinen, die in diese Stoffwechselwege involviert sind, in Bezug auf die Menge und/oder die Aktivität kann einen direkten Einfluß auf die Produktion oder die Effizienz der Produktion einer gewünschten Feinchemikalie haben. Beispielsweise kann eine Reaktion, die in direkter Konkurrenz zu einem Zwischenprodukt abläuft, welches für die gewünschte Feinchemikalie auftritt, ausgeschaltet, bzw. ein Stoffwechselweg, der für die Produktion dieses spezifischen Zwischenprodukts verantwortlich ist, optimiert werden.

Die Nutzung von Genen des Stoffwechsels zur Generierung von Mikroorganismen, vorzugsweise der Gattung *Ashbya*, insbesondere von *Ashbya gossypii* Stämmen, mit moduliertem Stoffwechsel ist noch nicht beschrieben.

Aufgabe der vorliegenden Erfindung ist daher die Bereitstellung neuer Targets zur Beeinflussung der Stoffwechselabläufe in Mikroorganismen der Gattung *Ashbya*, insbesondere in *Ashbya gossypii*. Insbesondere besteht die Aufgabe, bestimmte Stoffwechselwege in derartigen Mikroorganismen gezielt zu optimieren. Eine weitere Aufgabe ist die Verbesserung der Vitamin B2-Produktion durch derartige Mikroorganismen.

Gelöst wird obige Aufgabe 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, und zwar insbesondere:

a) eine, vorzugsweise niederregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer C1-Tetrahydrofolat-Synthase 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 72“ trägt.

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Be-

zeichnung „Oligo 72v“ trägt.

Ein erster 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: 4 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.

10

Die Inserts von „Oligo 72“ und „Oligo 72v“ besitzen mit dem MIPS Tag „Ade3“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 1 bzw. SEQ ID NO: 4. Die vom kodierenden Strang gemäß SEQ ID NO: 1 oder 4 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer C1-Tetrahydrofolat-Synthase aus *S. cerevisiae*.

15

b) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Argininosuccinat-Synthase kodiert.

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 81“ 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 81v“ trägt.

30

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 6. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 9 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 81“ und „Oligo 81v“ besitzen mit dem MIPS Tag „Arg1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 6

bzw. SEQ ID NO: 9. Die vom korrespondierenden Gegenstrang zu SEQ ID NO:6 bzw. vom kodierenden Strang gemäß SEQ ID NO:9 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Argininosuccinat-Synthase aus *S. cerevisiae*.

5

c) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Phosphoribosylamidoimidazol-Succinocarboxamid-Synthase 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 86“ 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 86v“ trägt.

20

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 12. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 14 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 86“ und „Oligo 86v“ besitzen mit dem MIPS Tag „Ade1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 12 bzw. SEQ ID NO: 14. Die vom korrespondierenden Gegenstrang zu SEQ ID NO:12 bzw. vom kodierenden Strang gemäß SEQ ID NO:14 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Phosphoribosylamidoimidazol-Succinocarboxamid-Synthase aus *S. cerevisiae*.

30

d) eine, vorzugsweise niederregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Fumarat-Reduktase kodiert.

35

Gemäß einer bevorzugten Ausführungsform dieses Aspekts der Erfindung wurde DNA-Klon isoliert, welcher für eine charakteristische Teilsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 162“ trägt.

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, welcher für die Vollsequenz der erfindungsgemäßen Nukleinsäuresequenz kodiert und die interne Bezeichnung „Oligo 162v“ trägt.

5

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 16. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 18 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.

10

Die Inserts von „Oligo 162“ und „Oligo 162v“ besitzen mit dem MIPS Tag „FRDS1 (2)“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 16 bzw. SEQ ID NO: 18. Die vom kodierenden Strang abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Fumarat-Reduktase aus *S. cerevisiae*.

15

e) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Phosphoenolpyruvat-Carboxykinase kodiert.

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 178“ 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 178v“ trägt.

30-

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 20. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 22 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 178 und „Oligo 178v“ besitzen mit dem MIPS Tag „PCK1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 20 bzw. SEQ ID NO: 22. Die vom korrespondierenden Gegenstrang zur SEQ ID NO: 20 bzw. vom kodierenden Strang gemäß SEQ ID NO: 22 abgeleiteten Aminosäuresequenzen besitzen signifikante Sequenzhomologie mit einer Phosphoenolpyruvat-Carboxykinase aus *S. cerevisiae*.

f) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Uroporphyrinogen-Decarboxylase 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 64“ 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 64v“ trägt.

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 24. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 26 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 64“ und „Oligo 64v“ besitzen mit dem MIPS Tag „Hem12“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 24 bzw. SEQ ID NO: 26. Die vom kodierenden Strang abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Uroporphyrinogen-Decarboxylase aus *S. cerevisiae*.

30

g) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Siroheme-Synthase 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 125“ trägt.

- 5 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 125v“ trägt.

- 10 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: 30 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.
- 15

- 20 Die Inserts von „Oligo 125“ und „Oligo 125v“ besitzen mit dem MIPS Tag „Met1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 28 bzw. SEQ ID NO: 30. Die vom Gegenstrang zu SEQ ID NO:28 bzw vom kodierenden Strang gemäß SEQ ID NO:30 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Siroheme-Synthase aus *S. cerevisiae*.

h) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Uroporphyrinogen-III-Synthase kodiert.

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 107“ 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 107v“ trägt.

- 35 Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 32. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 34 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 107“ und „Oligo 107v“ besitzen mit dem MIPS Tag „Hem4“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 32 bzw. SEQ ID NO: 34. Die vom kodierenden Strang abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Uroporphyrinogen-III-Synthase aus *S. cerevisiae*.

10

i) eine, vorzugsweise niederregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Phosphoglycerat-Kinase 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 136“ 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 136v“ trägt.

25

Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 36. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 38 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 Insert von „Oligo 136“ und „Oligo 136v“ besitzen mit dem MIPS Tag „Pgc1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 36 bzw. SEQ ID NO: 38. Die jeweils vom kodierenden Strang abgeleiteten Aminosäuresequenzen besitzen signifikante Sequenzhomologie mit einer Phosphoglycerat-Kinase aus *S. cerevisiae*.

35

k) eine, vorzugsweise niederregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion eines Proteinase-B-Inhibitors-2 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 157“ trägt.

- 5 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 157v“ trägt.

- 10 Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 40. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 42 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.
- 15

- 20 Die Inserts von „Oligo 157“ und „Oligo 157v“ besitzen mit dem MIPS Tag „PBI2“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 40 bzw. SEQ ID NO: 42. Die vom korrespondierenden Gegenstrang zur SEQ ID NO: 40 bzw. vom kodierenden Strang der SEQ ID NO: 42 abgeleiteten Aminosäuresequenzen besitzen signifikante Sequenzhomologie mit einem Proteinase-B-Inhibitor-2 aus *S. cerevisiae*.

- 25 I) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Cystein-Synthase 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 108“ 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 108v“ trägt.

- 40 Ein erster Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 44. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 47 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 108“ und „Oligo 108v“ besitzen mit dem MIPS Tag „CYSK“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 44 bzw. SEQ ID NO: 47. Die vom korrespondierenden Gegenstrang von SEQ ID NO:44 bzw. vom kodierenden Strang gemäß SEQ ID NO:47 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Cystein-Synthase aus *A. nidulans*.

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

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

- 20 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, 3, 5, 7, 8, 10, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 46, oder SEQ ID NO: 48 umfasst; sowie funktionale Äquivalente der erfindungsgemäßen Polypeptide bzw. Proteine.

- 25 Funktionale Äquivalente unterscheiden sich dabei von den erfindungsgemäß konkret offenbarten Produkten in ihrer Aminosäuresequenz durch Addition, Insertion, Substitution, Deletion oder 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.

- 35 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.

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
5 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*.

10

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).

15

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. Ein weiterer Gegenstand der Erfindung betrifft die Verwendung einer erfindungsgemäßen Ex-
20 pressionskassette, eines erfindungsgemäßen Vektors oder eines erfindungsgemäßen Wirts zur rekombinanten Herstellung eines erfindungsgemäßen Polypeptids nach obiger Definition.

25

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
30 Menge des mikrobiologisch produzierten Vitamins B2 und/oder des Präkursors und/oder eines 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.

35

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 mikrobiologi-

schen 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.

5

Als bevorzugte Beispiele für oben genannte Effektoren sind zu nennen:

- a) Antikörper oder antigenbindene Fragmenten davon;
 - b) von a) verschiedenen Polypeptid-Liganden, welche mit einem erfindungsgemäßen Polypeptid wechselwirken;
 - 10 c) niedermolekulare Effektoren, welche die biologische Aktivität eines erfindungsgemäßen Polypeptids modulieren;
 - d) Antisense-Nukleinsäuresequenzen, welche mit einer erfindungsgemäßen Nukleinsäuresequenz wechselwirken.
- 15 Oben genannte Effektoren mit Spezifität für wenigstens eines der oben definierten erfindungsgemäßen Targets 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.

25

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*.

30

Figurenbeschreibung:

Figur 1 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Strang von Position 1 bis 144 in SEQ ID NO:1) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Ade3 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

35

Figur 2 A zeigt ein weiteres Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 862 bis 551 in SEQ ID NO: 6) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Arg1 aus *S. cerevisiae* (untere Sequenz). Figur 2 B zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 912 bis 859 in SEQ ID NO:6) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Arg1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind jeweils zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 3 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 117 bis 1 in SEQ ID NO:12) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Ade1 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 Strang gemäß Position 1 bis 882 in SEQ ID NO:16) (obere Sequenz) und einer Teilsequenz des MIPS-Tags FRDS1 (2) 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 Gegenstrang zu Position 783 bis 1 in SEQ ID NO:20) (obere Sequenz) und einer Teilsequenz des MIPS-Tags PCK1 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 (mittlere Sequenz) und einer Teilsequenz des MIPS-Tags Hem12 aus *S. cerevisiae* (untere Sequenz). Über diesen beiden ist die Consensussequenz dargestellt. Positionen mit fehlender Homologie sind mit schwarzen Rechtecken symbolisiert.

Figur 7 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 964 bis 529 in SEQ ID NO:28) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Met1 aus *S. cerevisiae* (untere Sequenz). Identische Se-

quenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

5 Figur 8 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Strang in Position 107 bis 313 in SEQ ID NO:32) (obere Sequenz) und einer Teilsequenz des MIPS-Tags „Hem4“ aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

10 Figur 9 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Strang in Position 2 bis 91 in SEQ ID NO:36) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Pgc1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

15

Figur 10 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 713 bis 513 in SEQ ID NO:40) (obere Sequenz) und einer Teilsequenz des MIPS-Tags „PBI2“ aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

20

Figur 11 A zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 1596 bis 1459 in SEQ ID NO:44) (obere Sequenz) und einer Teilsequenz des MIPS-Tags CYSK aus *A. nidulans* (untere Sequenz). Figur 11 B zeigt
25 ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 1441 bis 971 in SEQ ID NO:44) (obere Sequenz) und einer Teilsequenz des MIPS-Tags CYSK aus *A. nidulans* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

30

Detaillierte Beschreibung der Erfindung:

Die erfindungsgemäßen Nukleinsäuremoleküle kodieren Polypeptide bzw. Proteine, die hier als Proteine des Stoffwechsels (z.B. mit Aktivität bezüglich der Biosynthese wenigstens eines ge-
35 wünschten Bioprodukts oder eines Zwischenprodukts davon oder dem Abbau von Nebenprodukten) oder kurz als „SW-Proteine“ bezeichnet werden. Diese SW-Proteine haben z.B. eine Funktion bei der Regulierung des Energiehaushalts eines lebenden Systems. Aufgrund der Ver-

fügbarekeit 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 Produzenten 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.

5

10

Die vorliegende Erfindung beruht auf der Bereitstellung neuer Moleküle, die hier als SW-Nukleinsäuren und SW-Proteine bezeichnet werden und am Stoffwechsel, insbesondere in *Ashbya gossypii*, (z.B. bei der Synthese oder Regulation von Stoffwechselenzymen) beteiligt sind. Die Aktivität der erfindungsgemäßen SW-Moleküle in *A. gossypii* beeinflusst die Vitamin B2-Produktion durch diesen Organismus. Vorzugsweise wird die Aktivität der erfindungsgemäßen SW-Moleküle so moduliert, dass die Stoffwechsel- und/oder Energiewege von *A. gossypii*, an denen die erfindungsgemäßen SW-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.

15

20

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.

25

Verbesserung der Vitamin B2-Produktion:

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

30

So kann durch die gezielte Steuerung des Stoffwechsels, die Effizienz der Produktion eines gewünschten Produkts (z.B. einer Feinchemikalie) erhöht bzw. gegenüber Konkurrenzprodukten optimiert werden. Auch können die Zellen robuster gegen äußere Einflüsse gemacht werden, so dass die Lebensfähigkeit und damit die Produktivität im Fermenter erhöht wird.

35

Die Mutagenese von einem oder mehreren erfindungsgemäßen SW-Proteinen kann auch zu SW-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 SW-Proteine Reaktionen, die in direkter Konkurrenz zu einem Zwischenprodukt der Zielverbindung ablaufen, ausschalten, bzw. den Stoffwechselweg, der für die Produktion dieses spezifischen Zwischenprodukts verantwortlich ist, unterbrechen, und dadurch die Produktion der gewünschten Zielsubstanz optimieren. Zu den beeinflussbaren Prozessen gehört neben der Biosynthese des Produkts auch der Aufbau der Zellwände, die Transkription, Translation, die Biosynthese von Verbindungen, die für das Wachstum und die Teilung von Zellen nötig sind (z.B. Nukleotide, Aminosäuren, Vitamine, Lipide usw.) (Lengeler et al. (1999)). Durch Verbesserung von Wachstum und Vermehrung dieser veränderten Zellen ist es möglich, die Lebensfähigkeit der Zellen in Kulturen im Großmaßstab zu steigern und auch ihre Teilungsrate so zu verbessern, dass eine vergleichsweise größere Anzahl von produzierenden Zellen in der Fermenterkultur überleben kann. Die Ausbeute, Produktion oder Effizienz der Produktion kann zumindest aufgrund der Anwesenheit einer größeren Anzahl lebensfähiger Zellen, die jeweils das gewünschte Produkt produzieren, 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.

„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.

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-

besondere auch dann gegeben, wenn die Reaktivitätsmuster zwischen Mutante und unverändertem Polypeptid qualitativ übereinstimmen, d.h. beispielsweise gleiche Substrate mit unterschiedlicher Geschwindigkeit umgesetzt werden.

5 „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
10 Beispiel Natrium-, Calcium-, Ammonium-, Eisen- und Zinksalze, sowie Salze mit organischen 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.

15

„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
20 oder mit einem primären oder sekundären Amin; N-Acylderivate freier Aminogruppen, hergestellt durch Umsetzung mit Acylgruppen; oder O-Acylderivate freier Hydroxylgruppen, hergestellt durch Umsetzung mit Acylgruppen.

25 „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.

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

35 „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. Sig-

nalpeptide, Enzyme, Immunoglobuline, Oberflächenantigene, Rezeptoren oder Rezeptorliganden.

- 5 Erfindungsgemäß mit umfasste „funktionale Äquivalente“ sind Homologe zu den konkret offenbarten Proteinen. Diese besitzen wenigstens 60 %, vorzugsweise wenigstens 75% ins besondere wenigsten 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.
- 10 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.
- 15 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.
- 20 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 Homologer aus einer degenerierten Oligonukleotidsequenz verwendet werden können.
- 25 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).
- 30
- 35 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
5 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.

10 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
15 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
20 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).

25 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:

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

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.

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

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.

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 offenbaren vollständigen Sequenzen oder ein Abschnitt davon als Hybridisierungs sonde und Standard-Hybridisierungstechniken (wie z.B. beschrieben in Sambrook, J., Fritsch, E.F. und 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 offenbaren 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 SA-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, vorzugs-

weise 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.

- 5 Weitere erfindungsgemäße Nukleinsäuresequenzen sind abgeleitet von SEQ ID NO: 1, 4, 6, 9, 12, 14, 16, 18, 20, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44 oder SEQ ID NO: 47 und unterscheiden sich davon durch Addition, Substitution, Insertion oder Deletion einzelner oder mehrerer Nukleotide, kodieren aber weiterhin für Polypeptide mit dem gewünschten Eigenschaftsprofil.
- 10 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
- 15 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.
- 20

- Weiterhin umfasst die Erfindung auch Nukleinsäuresequenzen, welchen mit oben genannten
- 25 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
- 30 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
- 35 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%, vor-

zugsweise 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 stringen-

5 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

10 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 be-

15

20 zeichneten 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-

25

30

35

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

5 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

10 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 Ver-

15 knüpfen 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.

20

In einer weiteren Ausführungsform ist das erfindungsgemäße Antisense-Nukleinsäuremolekül ein alpha-anomeres Nukleinsäuremolekül. Ein alpha-anomeres Nukleinsäuremolekül bildet spezifische doppelsträngige Hybride mit komplementärer RNA, wobei die Stränge im Gegensatz zu

25 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

30

35 der entsprechenden Nukleinsäure zu hemmen. Ein Ribozym mit Spezifität für eine erfindungsgemäße kodierende Nukleinsäure kann z.B. auf der Basis einer hierin konkret offenbarten 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
5 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
10 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).

15 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 Expres-
20 sionskonstrukte. 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 gegebe-
25 nenfalls 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 Se-
30 quenzen sind z.B. beschrieben in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990).

Zusätzlich zu den artifiziellen Regulationssequenzen kann die natürliche Regulationssequenz vor dem eigentlichen Strukturgen noch vorhanden sein. Durch genetische Veränderung kann
35 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 natürli-

che 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.

5

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 P₇P₁-Promotor. Prinzipiell können alle natürlichen Promotoren mit ihren Regulationssequenzen verwendet werden. Darüber hinaus können auch synthetische Promotoren vorteilhaft verwendet werden.

10

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.

15

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.

20

Die Herstellung einer Expressionskassette erfolgt durch Fusion eines geeigneten Promotors mit einer geeigneten erfindungsgemäßen 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 (1982) sowie in T.J. Silhavy, 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.

25

30

35

Das rekombinante Nukleinsäurekonstrukt bzw. Genkonstrukt wird zur Expression in einem geeigneten Wirtsorganismus vorteilhafterweise in einen wirtsspezifischen Vektor inseriert, der eine 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, 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:

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.

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).

Hefe-Expressionsvektor zur Expression in der Hefe *S. cerevisiae*, wie pYepSec1 (Baldari et al., (1987) Embo J. 6:229-234), pMF α (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 mar-

kers 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.

5 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, 10 Cold Spring Harbor, NY, 1989 beschrieben.

Rekombinante Mikroorganismen:

15 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, 20 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 25 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, - 30 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 Säugetier-, Hefe- oder Insektenquelle abgeleitet sein. Der zur homologen Rekombination verwendete Vektor kann alternativ derart ausgestaltet sein, daß das endogene Gen bei homologer Rekombi- 35 nation mutiert oder anderweitig verändert ist, jedoch noch das funktionelle Protein codiert (z.B. kann der stromaufwärts gelegene regulatorische Bereich derart verändert sein, daß dadurch die Expression des endogenen Proteins verändert wird). Der veränderte Abschnitt des SW-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.

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 Gattungen 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.

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, die ein entsprechendes Antibiotikaresistenzgen (z.B. G418 oder Hygromycin) tragen, lassen sich durch entsprechende Antibiotika-enthaltende Medien oder Nährböden selektieren. Markerproteine, die an der Zelloberfläche präsentiert werden, können zur Selektion mittels Affinitätschromatographie genutzt werden.

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 Säuger tierzellen, wie CHO-Zellen, und Vektoren, wie pcDNA3neo-Vektor, die für Säugetierzellen geeignet sind, zu verstehen.

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:

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.

- 5 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
10 (1989) beschrieben.

- 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,
15 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.

- 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,
20 Springer Verlag, New York, Heidelberg, Berlin beschrieben.

- 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
30 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
35 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.

- 10 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.
- 15 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.

20 Allgemeine experimentelle Angaben

a) Allgemeine Klonierungsverfahren

- 25 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.

30

b) Polymerasekettenreaktion (PCR)

PCR wurde nach Standardprotokoll mit folgendem Standardansatz durchgeführt:

- 35 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

Die Kultivierung von rekombinanten E. coli-Stämme DH5 α 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 Medium 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

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.

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

Die Überstandsfraktion 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.

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:**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.

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.

10 Beispiel 4:

Herstellung einer geordneten Genbank (CHIP-Technologie)

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.

Auf diese Weise wurde ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 72“ trägt und mit dem MIPS Tag „Ade3“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 1.

25

Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 81“ trägt und mit dem MIPS Tag „Arg1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 6.

30 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 86“ trägt und mit dem MIPS Tag „Ade1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 12.

35 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 162“ trägt und mit dem MIPS Tag „FRDS1 (2)“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 16.

Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 178“ trägt und mit dem MIPS Tag „AgPCK1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 20.

- 5 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 64“ trägt und mit dem MIPS Tag „Hem12“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 24.

- 10 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 125“ trägt und mit dem MIPS Tag „Met1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 28.

- 15 Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 107“ trägt und mit dem MIPS Tag „Hem4“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 32.

- 20 Auf diese Weise wurde ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 136“ trägt und mit dem MIPS Tag „Pgk1“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 36.

- Auf diese Weise wurde weiterhin ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 157“ trägt und mit dem MIPS Tag „PBI2“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 40.

- 25 Auf diese Weise wurde ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 108“ trägt und mit dem MIPS Tag „CYSK“ aus *A. nidulans* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 44.

30 **Beispiel 5:**

Auswertung der Sequenzdaten mit Hilfe einer BLASTX Suche

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

So wurden folgende Homologien mit einem Aminosäurefragment, überwiegend ausaus *S. cerevisiae*, ermittelt. Die entsprechenden Alignments sind in beiliegenden Figuren 1 bis 11 gezeigt.

5

a) Die vom kodierenden Strang von SEQ ID NO:1 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer C1-Tetrahydrofolat-Synthase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 1 bis 144 aus SEQ ID NO:1) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 1 dargestellt. Eine weitere Homologie wurde für die Aminosäure-Teilsequenz entsprechend den Nucleotiden 180 bis 287 in SEQ ID NO:1 festgestellt. SEQ ID NO: 2 und SEQ ID NO: 3 zeigen jeweils eine N-terminal verlängerte Aminosäure-Teilsequenz .

10

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer C1-Tetrahydrofolat-Synthase zugeordnet werden.

15

b) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:6 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Agininosuccinat-Synthase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Gegenstrang zu den Nucleotiden 862 bis 551 aus SEQ ID NO:6) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 2A dargestellt. Eine weitere davon abgeleiteten Aminosäure-Teisequenz (entsprechend den Gegenstrang zu den Nucleotiden 912 bis 859 aus SEQ ID NO:6) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 2B dargestellt. SEQ ID NO: 7 und SEQ ID NO: 8 zeigen jeweils eine N-terminal verlängerte Aminosäure-Teilsequenz .

20

25

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Agininosuccinat-Synthase zugeordnet werden.

c) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:12 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Phosphoribosylamidoimidazol-Succinocarboxamid-Synthase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 117 bis 1 aus SEQ ID NO:12) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 3 dargestellt. SEQ ID NO: 13 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

30

35

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Phosphoribosylamidoimidazol-Succinocarboxamid-Synthase zugeordnet werden.

- d) Die vom kodierenden Strang gemäß SEQ ID NO:16 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Fumarat-Reduktase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 1 bis 882 aus SEQ ID NO:16) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 4 dargestellt. SEQ ID NO: 17 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Fumarat-Reduktase zugeordnet werden.

10

- e) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:20 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Phosphoenolpyruvat-Carboxykinase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 783 bis 1 aus SEQ ID NO:20) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 5 dargestellt. SEQ ID NO: 21 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

15

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Phosphoenolpyruvat-Carboxykinase zugeordnet werden.

- f) Die vom kodierenden Strang zu SEQ ID NO:24 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Uroporphyrinogen-Decarboxylase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (SEQ ID NO: 25 entsprechend den Nucleotiden 441 bis 1058 aus SEQ ID NO:24) zeigt Homologie zu einer Teilsequenz des MIPS-Tags Hem12 aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 6 dargestellt.

25

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Uroporphyrinogen-Decarboxylase zugeordnet werden.

- g) Die vom kodierenden Gegenstrang zu SEQ ID NO:28 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Siroheme-Synthase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu den Nucleotiden 966 bis 529 aus SEQ ID NO:28) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 7 dargestellt. SEQ ID NO: 29 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

35

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Siroheme-Synthase bzw. einer Uroporphyrin-III-C-Methyltransferase zugeordnet werden.

h) Die vom kodierenden Strang zu SEQ ID NO:32 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Uroporphyrinogen-III-Synthase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 107 bis 313 aus SEQ ID NO:32) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 8 dargestellt. SEQ ID NO: 33 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Uroporphyrinogen-III-Synthase zugeordnet werden.

i) Die vom kodierenden Strang zu SEQ ID NO: 36 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Phosphoglycerat-Kinase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 2 bis 91 aus SEQ ID NO:36) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 9 dargestellt. SEQ ID NO: 37 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Phosphoglycerat-Kinase zugeordnet werden.

k) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:40 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem Proteinase-B-Inhibitor-2 aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 713 bis 513 aus SEQ ID NO:40) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 10 dargestellt. SEQ ID NO: 41 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion eines Proteinase-B-Inhibitors-2 zugeordnet werden.

l) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:44 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Cystein-Synthase aus *S. cerevisiae* und *A. nidulans*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 1596 bis 1459 aus SEQ ID NO:44) mit einer Teilsequenz des *A. nidulans* Enzyms ist in Figur 11A dargestellt. Eine weitere davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 1441 bis 971 aus SEQ ID NO:44) mit einer Teilsequenz des *A. nidulans* Enzyms ist in Figur 11B dargestellt. SEQ ID NO: 45 und SEQ ID NO: 46 zeigen jeweils eine N-terminal verlängerte Aminosäure-Teilsequenz.

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit der Funktion einer Cystein-Synthase zugeordnet werden.

5 Beispiel 6:

Isolierung der Full-Length-DNA

a) Konstruktion einer *A. gossypii*-Genbank

- 10 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
- 15 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.
- 20 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
- 25 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
- 30 (*recA13*, *araD139*, Δ (*ara*,*leu*)7696, Δ (*lac*)17A, *galU*, *galK*, *hsrR*, *rps*(*str*^r), *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

35

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

K₂HPO₄/13,2 mM KH₂PO₄, 1,7 mM Natriumcitrat, 0,4 mM MgSO₄, 6,8 mM (NH₄)₂SO₄, 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 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 dupli-
5 ziert. 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 Mengen an Flüssigkeit aus 96 Vertiefungen einer Mikrotiterplatte auf Nylonmembran (GeneSc-
10 reen *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 Zellkonfluenz wurden die Membranen, wie von Herrmann, B. G., Barlow, D. P. und Lehrach, H.
15 (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.

Mit Hilfe des Random-Hexamer-Primer-Verfahrens (Feinberg, A. P. und Vogelstein, B. (1983);
20 Anal. Biochem. 132, S. 6-13) wurden doppelsträngige Sonden durch Aufnahme von [alpha-³²P]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 ³²P-markierten Sonden (0,5-1 x 10⁶cpm/ml) hybridisiert. Typischerweise wurden
25 Waschschriffe 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 bei 95 °C für 2 x 20 min in 2 mM Tris-HCl, pH 8,0, 0,2 mM EDTA, 0,1 % SDS gestrippt.

30 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
35 µg/ml) ausgestrichen. Einzelne Kolonien wurden danach zur Inokulierung von Flüssigkulturen für Herstellung von DNA mittels Alkali-Lyse-Verfahren (Birboim, H. C. und Doly, J. (1979), Nucleic Acids Res. 7, S. 1513-1523) verwendet.

d) Full-Length DNA

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

„Oligo 72v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 4.

10 „Oligo 81v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 9. Das davon kodierte Protein umfasst vorzugsweise wenigstens eine der Aminosäuresequenzen gemäß SEQ ID NO: 10 und 11.

15 „Oligo 86v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 14.

„Oligo 162v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 18.

20 „Oligo 178v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 22.

„Oligo 64v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 26.

25 „Oligo 125v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 30.

30 „Oligo 107v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 34.

„Oligo 136v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 38.

35 „Oligo 157v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 42.

„Oligo 108v“. Das die Vollsequenz umfassende Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 47.

Tabelle 1: Sequenzübersicht

5

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
1	072	DNA-Teilsequenz	C1-Tetrahydrofolat-Synthase aus <i>S. cerevisiae</i>
2	072	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 1	
3	072	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 1	
4	072	DNA-Full-lenght-Sequenz	
5	072	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 4 von Position 440 bis 3253	
6	081	DNA-Teilsequenz	Agininosuccinat-Synthase aus <i>S. cerevisiae</i>
7	081	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 6	
8	081	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 6	
9	081	DNA-Full-lenght-Sequenz	
10	081	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 9 von Position 716 bis 1957	
11	081	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 9 von Position 2246 bis 2839	Phosphoribosylamidoimidazol- Succinocarboxamid-Synthase aus <i>S. cerevisiae</i>
12	086	DNA-Teilsequenz	
13	086	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 12	
14	086	DNA-Full-lenght-Sequenz	

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
15	086	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 14 von Position 587 bis 1492	
16	162	DNA-Teilsequenz	Fumarat-Reduktase aus <i>S. cerevisiae</i>
17	162	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 16	
18	162	DNA-Full-lenght-Sequenz	
19	162	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 18 von Position 264 bis 1772	
20	178	DNA-Teilsequenz	Phosphoenolpyruvat-Carboxykinase aus <i>S. cerevisiae</i>
21	178	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 20	
22	178	DNA-Full-lenght-Sequenz	
23	178	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 22 von Position 394 bis 2016	
24	64	DNA-Teilsequenz	Uroporphyrinogen-Decarboxylase aus <i>S. cerevisiae</i>
25	64	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 24	
26	64	DNA-Full-lenght-Sequenz	
27	64	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 26 von Position 612 bis 1712	
28	125	DNA-Teilsequenz	Siroheme-Synthase aus <i>S. cerevisiae</i>
29	125	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 28	
30	125	DNA-Full-lenght-Sequenz	

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
31	125	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 30 von Position 275 bis 2023	
32	107	DNA-Teilsequenz	Uroporphyrinogen-III-Synthase aus <i>S. cerevisiae</i>
33	107	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 32	
34	107	DNA-Full-lenght-Sequenz	
35	107	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 34 von Position 250 bis 1035	
36	136	DNA-Teilsequenz	Phosphoglycerat-Kinase aus <i>S. cerevisiae</i>
37	136	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang zu SEQ ID NO: 36	
38	136	DNA-Full-lenght-Sequenz	
39	136	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 38 von Position 447 bis 1790	
40	157	DNA-Teilsequenz	Proteinase-B-Inhibitor-2 aus <i>S. cerevisiae</i>
41	157	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 40	
42	157	DNA-Full-lenght-Sequenz	
43	157	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 42 von Position 809 bis 1033	
44	108	DNA-Teilsequenz	Cystein-Synthase aus <i>S. cerevisiae</i> und <i>A. nidulans</i>
45	108	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 44	
46	108	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 44	

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
47	108	DNA-Full-lenght-Sequenz	
48	108	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 47 von Position 450 bis 1517	

Patentansprüche

1. Polynukleotid, welches aus *Ashbya gossypii* isolierbar ist und für ein Protein kodiert, das mit dem Stoffwechsel von *A. gossypii* assoziiert ist.
5
2. Polynukleotid, nach Anspruch 1, das mit dem Stoffwechsel von *A. gossypii* assoziiert ist und eine in Tabelle 1 angegebene strukturelle und/oder funktionale Eigenschaft besitzt.
3. Polynukleotid nach Anspruch 1 oder 2, umfassend eine Nukleinsäuresequenz gemäß
10 SEQ ID NO: 1, 6, 12, 16, 20, 24, 28, 32, 36, 40 oder 44, 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 3, welches eine Nukleinsäuresequenz gemäß SEQ ID NO: 4, 9, 14, 18, 22, 26, 30, 34, 38, 42, oder 47 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 eine Nukleinsäuresequenz gemäß einem der Ansprüche 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
30 Aminosäurereste gemäß SEQ ID NO: 2, 3, 5, 7, 8, 10, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 46, oder SEQ ID NO: 48 umfasst; sowie funktionale Äquivalente davon, insbesondere solche welche die Aktivität vergleichbar mit derjenigen eines Proteins aus *S. cerevisiae*, *A. nidulans* oder *A. gossypii* mit einer in Tabelle 1 angegebenen Aktivität besitzen.
35
8. Expressionskassette, umfassend in operativer Verknüpfung mit wenigstens einer regulativen Nukleinsäuresequenz eine Nukleinsäuresequenz nach einem der Ansprüche 1 bis 6.

9. Rekombinanter Vektor, umfassend wenigstens eine Expressionskassette nach Anspruch 8.
- 5 10. Prokaryotischer oder eukaryotischer Wirt, transformiert mit wenigstens einem Vektor nach Anspruch 9.
11. Prokaryotischer oder eukaryotischer Wirt, in welchem die funktionale Expression wenigstens eines Gens moduliert ist, das für ein Polypeptid nach Anspruch 7 kodiert;
10 oder in welchem die biologische Aktivität eines Polypeptids nach Anspruch 7 erniedrigt oder erhöht ist.
12. Wirt nach Anspruch 10 oder 11 aus der Gattung *Ashbya*.
- 15 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.
- 20 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.
- 25 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
30 nach Anspruch 7 oder einer 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.
- 35 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.

- 5 17. Effektor für ein Target, ausgewählt unter einem Polypeptid nach Anspruch 7 oder einer dafür kodierenden Nukleinsäuresequenz, wobei der Effektor ausgewählt ist unter:
- a) Antikörpern oder antigenbindenden Fragmenten davon;
 - b) von a) verschiedenen Polypeptid-Liganden, welche mit dem Polypeptid gemäß Anspruch 5 wechselwirken;
 - 10 c) niedermolekularen Effektoren, welche die biologische Aktivität eines Polypeptids nach Anspruch 5 modulieren;
 - d) Antisense-Nukleinsäuresequenzen.
- 15 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.
- 20 19. Verfahren nach Anspruch 18, wobei man vor und/oder während der Kultivierung des Wirts diesen mit einem Effektor nach Anspruch 17 behandelt.
- 25 20. Verfahren nach Anspruch 18 oder 19, wobei der Wirt ausgewählt ist unter Mikroorganismen der Gattung *Ashbya*.
- 30 21. Verfahren nach einem der Ansprüche 18 bis 20, wobei der Mikroorganismus ein Wirt nach einem der Ansprüche 10 bis 12 ist.
22. Verwendung eines Polynukleotids nach einem der Ansprüche 1, 2 und 4 oder eines Polypeptids nach Anspruch 5 als Target zur Modulation der Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon in einem Mikroorganismus der Gattung *Ashbya*.
- 35 23. Verwendung eines Polynukleotids nach einem der Ansprüche 1 bis 4 und 6 oder eines Polypeptids nach Anspruch 7 als Target zur Modulation des einer Stoffwechselfunktion in einem Mikroorganismus der Gattung *Ashbya* während der

Kultivierung zur mikrobiologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon.

24. Wirt nach Anspruch 12 mit verbesserter Stoffwechselregulation oder modifizierter Stoffwechselfunktion.
- 5

1/11

Fig. 1**Oligo 72**

1	DLSP	TAQ	KKI	ALY	TAQ	GFD	KLP	ICIA	KTQ	YSL	SHD	ATF	EER	PLR	FC	VP	144	<i>A. gossypii</i>
	D+	P	AQ	+KI	+Y	QGF	LP	ICIA	KTQ	YSL	SHD	AT	+	P	F	P		
852	DIL	PEA	QR	KID	MY	KEQ	GFG	NLP	ICIA	KTQ	YSL	SHD	ATL	KGV	PTG	FTFP	899	<i>S. cerevisiae</i>

2/11

Fig. 2A**Oligo 81**

862 YSRLTLASSTMAPTFRLRNAEYIRSMIKPSQITVNGQVRLRLYKGNVIVLGRSSETENLYD
YSRL + P EYIRSMI+PSQ +VNG VR+RLYKGNVI+LGRS++TE LYD
321 YSRLIYNGFLLHP----ECEYIRSMIQPSQNSVNGTVRVRLYKGNVIILGRSTKTEKLYD 376

ATESSMDELTGFGPTD TTGFIAIQAIRMKKYGQSKDKGESLTL 551 A. gossypii
TESSMDELTGF PTD TTGFIAIQAIR+KKYG+SKK KGE LTL
377 PTESSMDELTGFLPTD TTGFIAIQAIRIKKYGESKKTKEELTL 420 S. cerevisiae

Fig. 2B

912 EGLTLDKEVRHLRDQFVT 859 A. gossypii
EGLTLDKEVR LRD FVT
301 EGLTLDKEVRQLRDSFVT 318 S. cerevisiae

3/11

Fig. 3**Oligo 86**

```
117 MSIVQTQLGGILPLVARGKVRDIYEVDQETLLFVATDRI 1  A. gossypii
      MSI QT L  ILPLVARGKVRDIYEVD++TLLFVATDRI
1    MSITKTELGNILPLVARGKVRDIYEVDEKTLLFVATDRI 39  S. cerevisiae
```

4/11

Fig. 4

Oligo 162

1 DLELADDAVQGLRYMDKAGAHILHTKNVVFCTGGFGSSKELLGKYAPHLRDLPTTNAQG 180
D+ D ++ + Y DK G +H++ +VVFC+GGFG SKE+L +YAP L +LPTTN Q
165 DIHEKDGSISAVVYEDKNGEKHMVSANDVVFCSGGFGFSKEMLKKEYAPELVNLPPTTNGQQ 224

181 TTGDGQDLMLRVGGQLIDMEHIQVHPTSFVDPKDKDSGSKFLAAEALRGNGGVLLNPSTG 360
TTGDGQ L+ ++G LIDM+ IQVHPT F+DP D+ S KFLAAE+LRG GG+LLNP TG
225 TTGDGQRLQLKLGADLIDMDQIQVHPTGFIDPNDRSSWKFLAAESLRGLGGILLNPITG 284

361 RRFTNELATRDVLTASIQEHC--KEHVAYLVLSQGAYETLQSFVNFYISKGLMRASTVGE 534
+RF NEL TRDV+TA+IQ+ C +++ A LV+ + Y L++ ++FY+ K L++ T+ +
285 KRFVNELTTRDVVTAAIQKVCPOEDNRALLVMGEKMYTDLKNNLDFYMFKKLVQKLTLTSLQ 344

535 LSDEIGVP----KQXXXXXXXXXXXXXVDQFQRRHIINDFGPSVDGSTAIYVGLITPAVH 702
+ E +P + D R I+N+FG V T +++G +TP VH
345 VVSEYNLPITVAQLCEELQTYSSFTTKADPLGRTVILNEFGSDVTPETVVFIGEVTPVH 404

703 FTMGGVRINSRAEVL-SSAGTTYKGLYAAGEVSGGVHGRNRLGGSSLLECVVFGRQAATS 879
FTMGG RIN +A+V+ + KGLYAAGEVSGGVHG NRLGGSSLLECVVFGR AA S
405 FTMGGARINVKAQVIGKNDERLLKGLYAAGEVSGGVHGANRLGGSSLLECVVFGRTAES 464

880 I 882 A. gossypii
I
465 I 465 S. cerevisiae

5/11

Fig. 5

Oligo 178

783 MSPSKVYQTP-----EEKIRAEGLSSGVLTIIRNAPAAVLYEDALAERKTAISSTGAL 622
MSPSK+ T E+KIR EL LS V TIRRNAPAAVLYED L E KT ISS+GAL
1 MSPSKMNATVGSTSEVEQKIRQELALSDEVTTIRRNAPAAVLYEDGLKENKTVISSSGAL 60

621 IAYSGEKTGRSPKDKRIVDEPTSSDNIWWGPVNRKASEKTWLINRERAADFLRTREHLYI 442
IAYSG KTRSPKDKRIV+EPTS D IWWGPVN+ SE+TW INRERAAD+LRTR+H+YI
61 IAYSGVKTGRSPKDKRIVEEPTSKDEIWWGPVNKPCERTWSINRERAADYLRTRDHIYI 120

441 VDAYAGWDPYRIKIRIVCCRAYHALFMTNMLIRPTEEEELAEFGEPDFTVWNAGQFPANT 262
VDA+AGWDP+YRIK+R+VC RAYHALFMTNMLIRPTEEEELA FGEPDFTVWNAGQFPAN
121 VDAFAGWDPKYRIKVRVVCARAYHALFMTNMLIRPTEEEELAHFGEPDFTVWNAGQFPANL 180

261 HTEGMSSKTTVEINFRAMEMVILGTEYAGEMKKGIFTVMFYLMPIHNVLTLHSSANQG- 85
HT+ MSSK+T+EINF+AMEM+ILGTEYAGEMKKGIFTVMFYLMP++HNVLTLHSSANQG
181 HTQDMSSKSTIEINFKAMEMIILGTEYAGEMKKGIFTVMFYLMPVHHNVLTLLHSSANQGI 240

84 PNNDVTLLFFGLSGTGKTTLSADQHRKLI 1 A. gossypii
N DVTLLFFGLSGTGKTTLSAD HR LI
241 QNGDVTLLFFGLSGTGKTTLSADPHRLLI 268 S. cerevisiae

6/11

Fig. 6

Oligo 64

Consensus	F KNDLILRAAKGE VERPPCWIMRQAGRYL
A. gossypii	F ■■■ KNDLMLRAARGE ■ VERPPCWIMRQAGRYL
S. cerevisiae	F ■■■ KNDLILRAAKGE ■ VERPPCWIMRQAGRYL
	PEYHDVK RDFF TCRDAEIASSEITI QPVRRY
	PEYHDVK ■■ RDFF ■ TCRDAEIASSEITL QPVRRY
	PEYHEVK ■■ RDFF ■ TCRDAEIASSEITI QPVRRY
	GLIDAAIIFSDILVIPQAMGMRVEMLEGKGPH
	■ GLIDAAIIFSDILVIPQAMGMRVEMLEGKGPH
	■ GLIDAAIIFSDILVIPQAMGMRVEMLEGKGPH
	FPEPLR E L VLDY VDV LKELDWAFKAIT
	FPEPLR ■■ E ■ L ■■ VLEY ■ VDV LRELDWAFRAIT
	FPEPLR ■■ E ■ L ■■ VLDY ■ VDV LKELDWAFKAIT
	LTR KL GDVPLFGFCGGPWTLLVYMTEGGGSR
	LTR ■ KL ■ GDVPLFGFCGGPWTLLVYMTEGGGSR
	MTR ■ KL ■ GEVPLFGFCGGPWTLMVYMTEGGGSR
	LFRFAK WI PELAHKLLQKITDVAIEFLAQ
	LFRFAK ■ WI ■■■ PELAHKLLQKITDVAIEFLAQ
	LFRFAK ■ WI ■■■ PELSHKLLQKITDVAVEFLSQ
	QV AG QILQLFESWGGELS D DEFSLPYLK
	QV ■ AG ■ QILQLFESWGGELS ■■ D ■ DEFSLPYLK
	QV ■ AG ■ QILQVFESWGGELS ■■ D ■ DEFSLPYLR
	QIA VP RL ELGI E IPM VFAKGSWYALD
	QIA ■■ VP ■ RL ■ ELGI ■ E ■ VPM ■ VFAKGSWYALD
	QIA ■■ VP ■ RL ■ ELGI ■ E ■ IPM ■ VFAKGSWYALD
	KLC AGF VSLDW WDP EAVKIN RVTLQG
	KLC ■ AGY ■■ VSLDW ■ WDP ■ EAVRIN ■■ RVTLQG
	KLC ■ SGF ■■ VSLDW ■ WDP ■ EAVKIN ■■ RVTLQG
	NLDP VMYGS E ITKK MI AFGGGK YI
	NLDP ■ VMYGS ■ E ■ ITKK ■■ MI ■ GFGGGK ■■ YI
	NLDP ■ VMYGS ■ E ■ ITKK ■■ MI ■ AFGGGK ■■ YI
	VNFGHGTHPFMDPD IK FLEECHRIG
	VNFGHGTHPFMDPE ■ IR ■ FLEECHRIG
	VNFGHGTHPFMDPD ■ IK ■ FLEECHRIG

Consensus
A. gossypii
S. cerevisiae

7/11

Fig. 7

Oligo 125

964 YIFGRGGEEYLLFKRHGFEPVVPGISSALAATTVAHIPATQQ*RRRAGLDMYQAPGRRG
YIFGRGGEE+ FK HG+ P V+PGISS+LA T +A IPATQ+ L + GR+G
415 YIFGRGGEEFNFFKDHGYIPVVLPGISSSLACTVLAQIPATQRDIADQVL-ICTGTGRKG 473

VLPPAPEFVAARTTIFLMALHRVGDIVQALLEQQWDPQVPAAIVERASCPDQRVTRTTLA
LP PEFV +RTT+FLMALHR L+ LL+ WD VPAAIVER SCPDQRVTRT L
474 ALPIIPEFVESRTTVFLMALHRANVLITGLLKHGWDGDVPAAIVERGSCPDQRVTRTLLK 533

SVPEVVEQIGSRPPGLLIVGHAVTSL 529 A. gossypii
VPEVVE+IGSRPPG+L+VG AV +L
534 WVPEVVEEIGSRPPGVLVVGKAVNAL 559 S. cerevisiae

8/11

Fig. 8**Oligo 107**

107 ERVFTDNSWVVLFPQGTNELLP LLRQGAA-----KIAAIGPTTAQYLLDNATPPILTS 268
+ VF+D WVV+FSPQGT E+ L ++A+IGPTT +YL DN + S
199 DEVFSD--WVVVFSPQGTKEITQYLGDSNRLPGSHLRVASIGPTTKKYLLDDNDVTSDVVS 256

269 PKPEPSSLYNAIKDY 313 A. gossypii
PKP+P SL +AI+ Y
257 PKPDPKSLLD AIELY 271 S. cerevisiae

9/11

Fig. 9**Oligo 136**

2 ISHVSTGGGASLELLEGGKDLPGVTFLSSKQ 91 A. gossypii
ISHVSTGGGASLELLEGGK+LPGVTFLS+KQ
387 ISHVSTGGGASLELLEGGKELPGVTFLSNKQ 416 S. cerevisiae

10/11

Fig. 10

Oligo 157

713 SYIITLKKGA---DLAKVRESIEQIGGSILHEHSLINGLSVKLPEVTALEQIKSQHDDLI 543
++I+TLKK + K +S+ GGSILHE +I G ++K+P+V L ++K +H+D+I
3 NFIVTLKKNTPDVEAKKFLDSVHHAGGSILHEFDIIKGYTIKVPDVLHLNKLKEKHNDVI 62

542 QHIEKDQEMH 513 A. gossypii
+++E+D+E+H
63 ENVEEDKEVH 72 S. cerevisiae

11/11

Oligo 108**Fig. 11A**

1596 HAAALDNAVWTNQFDNEANWKAHYATTGPEILQQMRNAGLEVDAFT 1459 A. gossypii
 HA +LDNAVWTNQFDN AN +AH TTGPEI Q G ++DAFT
 164 HAESLDNAVWTNQFDNTANRRAHIE TTGPEIWAQ---TGGKLDAFT 206 A. nidulans

Fig. 11B

1441 GGTFSGIARYLAEATERKCKLVVTDPPGSVVYSFVKTN--LREGSSFTEGVGQGRITK 1268
 GGT +GI YL +A+ + K + DPPGSV++S++++G +ER GSS TEG+GQGRIT
 212 GGTLAGITYYLLKQASGGRVKSFLADPPGSVLHSYIQSGGKLVERSGSSITEGIGQGRITD 271

1267 NLEASVDIIDDAVFVPDPESLNMTYRLLDEEGLFLGGTSGLVAGAVRVARELPPGSNVV 1088
 NL+ V +D ++ + D +++ M YR LDEEGL+LG +S LNV A VA +L GS VV
 272 NLQPDVGTLDGSLNISDEKTIEMIYRCLDEEGLYLGAASSALNVVAAKEVAEKLKGSTVV 331

1087 TILCDTGHKYADRVFSKKWLQEKKLYDEIPDSLKKYVTL 971 A. gossypii
 TIL D ++YADR+FSK WL+ K L + IP L+KY+ L
 332 TILADGAYRYADRLFSKSWLESKGLRNAIPKHLEKYIVL 370 A. nidulans

SEQUENCE LISTING

<110> BASF Aktiengesellschaft

<120> Neue Polynukleotide zur Modulation von Stoffwechselreaktionen

<130> M/42301/PCT

<160> 48

<170> PatentIn version 3.1

<210> 1

<211> 1058

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 72

<400> 1
gatctctctc ctacggcgca aaagaagatt gccttgtaga cggcccaagg ttttgataaa 60
ctacctatct gtatcgcgaa aactcagtag tccctgtccc atgatgccac ctttgaagaa 120
cgtccctctc gatcttggtg tccccataag ggatatcagg gccttcatt ggagcgggct 180
acctgtacgc gctagcggca gagattcaaa ctatccccgg tctctccact aacgcaggct 240
acatgaatgt tgaagtcagt gatgacggcg agattgaagg tcttttctag gtctacgatt 300
acgtatgctt agccaccccc gctcatatac ataattatgt tgaacaggca tgctcaggct 360
gcgttccctg tagtgacccc gttagtgcgc atgggccccg aagccgttgc ttttagaacc 420
ggcatgcgac acatacggga gacattactt tagcgtcgca gtgacagtta tgtcccgaga 480
cgctggcatg caatcgatgt cgagcctagt gagtcagaaa cggagtcagt gtcttcagac 540

```

acacaagcga aatcaggggtc agcttttgact aaacctaata accatagtgc attacgtccg      600
actgtatctt cacataaagt tagggaggag aatcaaagca aggtagcata acaaagtatg      660
gatttgatgg tcataaataa ggacgactgg gtatatgtac agtccgaggg cgaccacagc      720
ggcgatcaca cttattttga gagcacaatc acctcatcgc ttcgagacac aaagcggcac      780
gcgctattcc cgcccatagt cagggagcag ttcatccagg ggaaagatat gtgcctacca      840
atctaccacg acgggataaa agagcaggca gcgctcgagt acgacggcac aggcgcgcta      900
tgccggccttt gacacgctgt ggcaccacgc cgccacacag cacgcaataa acgggcaacg      960
gtatattgct gatatacctc gatacttatt attagaaaaa acctgcggtt gcaggctttt     1020
atactacata aagacaggat gaaaaacgta gatagatc                               1058

```

<210> 2

<211> 48

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 72

<400> 2

```

Asp Leu Ser Pro Thr Ala Gln Lys Lys Ile Ala Leu Tyr Thr Ala Gln
1           5           10          15

```

```

Gly Phe Asp Lys Leu Pro Ile Cys Ile Ala Lys Thr Gln Tyr Ser Leu
          20          25          30

```

```

Ser His Asp Ala Thr Phe Glu Glu Arg Pro Leu Arg Phe Cys Val Pro
          35          40          45

```

<210> 3

<211> 41

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 72

<400> 3

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

Ile Pro Gly Leu Ser Thr Asn Ala Gly Tyr Met Asn Val Glu Val Ser
20 25 30

Asp Asp Gly Glu Ile Glu Gly Leu Phe
35 40

<210> 4

<211> 4474

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (440)..(3253)

<223>

<220>

<221> misc_feature

<223> Oligo 72

<400> 4

cggtttcagt tttttttatc cccaattact ctggttactg attccgcgat tgacgctaatt 60
ttaacttttag cttttttatat gcttgactta gcctttacct gttgtcgact ttcttttagtt 120
cagcatcttg tatattactt gttcacgtat aaatagatag cgatacccct ttgcgagtat 180
acagttgcct tactgtgatt atgtttcttc tggttttatct gctaggtaag ttttgacctt 240
tttttgatgt tcacaagcac cgtcgacgct caagaaaacg atcgaagttc ctccctaacg 300
tctgccgttt accaagtata acgttgaact gtactatcct cctgtgcact aagtttagcat 360

gcttcgcacc tttacggagt ggagtattct aaggagtaat actcagctcg gggctttcag 420

ctgcccagta cgaagacat tac cat tta att tct ggt aaa agc atc gcg aaa 472
 Tyr His Leu Ile Ser Gly Lys Ser Ile Ala Lys
 1 5 10

acc att aga gac aat gct ctg aga gag gtg aaa gag att agg cag caa 520
 Thr Ile Arg Asp Asn Ala Leu Arg Glu Val Lys Glu Ile Arg Gln Gln
 15 20 25

gtt cca gat ttt aat cca aaa ctg aag atc ttg cag gtt ggc tcc cgc 568
 Val Pro Asp Phe Asn Pro Lys Leu Lys Ile Leu Gln Val Gly Ser Arg
 30 35 40

cct gat tca tct gca tat gtc cgg atg aag ttg aaa gca tcc acc gag 616
 Pro Asp Ser Ser Ala Tyr Val Arg Met Lys Leu Lys Ala Ser Thr Glu
 45 50 55

tct ggt gta gaa tgt gac gtt gaa aag ctt cca gaa gat att acg caa 664
 Ser Gly Val Glu Cys Asp Val Glu Lys Leu Pro Glu Asp Ile Thr Gln
 60 65 70 75

ctc gaa cta ctc aga aag atc gag aaa atc aac tct gac gcc gaa gtg 712
 Leu Glu Leu Leu Arg Lys Ile Glu Lys Ile Asn Ser Asp Ala Glu Val
 80 85 90

cat ggg att ctg att cag ctg cca ctg cca aag cat tta gat gaa gtg 760
 His Gly Ile Leu Ile Gln Leu Pro Leu Pro Lys His Leu Asp Glu Val
 95 100 105

aca gtt acc aat gct gtg gac cat gaa aag gac gtt gat ggg ttc cat 808
 Thr Val Thr Asn Ala Val Asp His Glu Lys Asp Val Asp Gly Phe His
 110 115 120

cgc tat aat gtt ggt gag ctc gct aaa cgt ggc ggg aaa cca tac ttc 856
 Arg Tyr Asn Val Gly Glu Leu Ala Lys Arg Gly Gly Lys Pro Tyr Phe
 125 130 135

ctg cca tgt aca cct aat gcg tgt ata cag ctt ttg aaa gag tgc ggt 904
 Leu Pro Cys Thr Pro Asn Ala Cys Ile Gln Leu Leu Lys Glu Cys Gly
 140 145 150 155

gtt gag ata aga ggg aaa cag gca gtc gta att ggc cgc tct gac atc 952
 Val Glu Ile Arg Gly Lys Gln Ala Val Val Ile Gly Arg Ser Asp Ile
 160 165 170

gtg ggg acc cca gtt gca acg atg ctc aaa aat ctt gat gcc acg gtg 1000
 Val Gly Thr Pro Val Ala Thr Met Leu Lys Asn Leu Asp Ala Thr Val
 175 180 185

act gtt gcg cac aga cat acg aca aat ctt cca caa gtc ctc tcc act 1048
 Thr Val Ala His Arg His Thr Thr Asn Leu Pro Gln Val Leu Ser Thr
 190 195 200

gcg gat att gta gtt gct gcg tgt ggc gtg cct aac tac ata aag ggc 1096
 Ala Asp Ile Val Val Ala Ala Cys Gly Val Pro Asn Tyr Ile Lys Gly
 205 210 215

gaa tgg tta aaa gat ggc gcc gtg gtt att gac gtt ggc atc aat tat 1144
 Glu Trp Leu Lys Asp Gly Ala Val Val Ile Asp Val Gly Ile Asn Tyr
 220 225 230 235

gtt cag gat tcc acc aaa aaa tca ggc caa aga tta gtc ggt gat gtg Val Gln Asp Ser Thr Lys Lys Ser Gly Gln Arg Leu Val Gly Asp Val 240 245 250	1192
gac ttc gaa tcg gca aag gag aag gct tct cac atc acg cct gta cct Asp Phe Glu Ser Ala Lys Glu Lys Ala Ser His Ile Thr Pro Val Pro 255 260 265	1240
ggc ggc gtg ggt cct atg acc gtg gcc atg ctg gtg caa aat gtc ttg Gly Gly Val Gly Pro Met Thr Val Ala Met Leu Val Gln Asn Val Leu 270 275 280	1288
act gca gca aaa aga tcg ttg cag aag gac gca gag gtg cca gtg atc Thr Ala Ala Lys Arg Ser Leu Gln Lys Asp Ala Glu Val Pro Val Ile 285 290 295	1336
aca cca ctg cct gtg aat ttt gag atg cct gtg cca tcg gat atc gat Thr Pro Leu Pro Val Asn Phe Glu Met Pro Val Pro Ser Asp Ile Asp 300 305 310 315	1384
ata tca agg aaa cag aat ccg aaa cat att tct cag gtg gca gcc gag Ile Ser Arg Lys Gln Asn Pro Lys His Ile Ser Gln Val Ala Ala Glu 320 325 330	1432
ttg ggt att agg gat cac gag ctg gag ttg ttt ggg cat tat aaa gcc Leu Gly Ile Arg Asp His Glu Leu Glu Leu Phe Gly His Tyr Lys Ala 335 340 345	1480
aag att tct cca aaa ctc ctc caa aga ttg cat gct aac caa aat gca Lys Ile Ser Pro Lys Leu Leu Gln Arg Leu His Ala Asn Gln Asn Ala 350 355 360	1528
aac tat gtc tta gtt gct ggg att aca ccg acc cct cta gga gag gga Asn Tyr Val Leu Val Ala Gly Ile Thr Pro Thr Pro Leu Gly Glu Gly 365 370 375	1576
aag tcg acc acc act atg ggg cta gtc caa gcg ctt tcc gct cac cta Lys Ser Thr Thr Thr Met Gly Leu Val Gln Ala Leu Ser Ala His Leu 380 385 390 395	1624
ggc aaa cgg gct att gca aac gtt cgt caa cct tca atg gga ccc acg Gly Lys Arg Ala Ile Ala Asn Val Arg Gln Pro Ser Met Gly Pro Thr 400 405 410	1672
ttt gga gtc aag gga ggc gcg gca ggc ggg ggt tac gcc cag gtc att Phe Gly Val Lys Gly Gly Ala Ala Gly Gly Gly Tyr Ala Gln Val Ile 415 420 425	1720
cca atg gat gag ttc aac ttg cat ttg aca ggc gac atc cac gcc att Pro Met Asp Glu Phe Asn Leu His Leu Thr Gly Asp Ile His Ala Ile 430 435 440	1768
tca gcc gcc aac aat ctg cta gcc gca gcc ata gat aca agg atg ttt Ser Ala Ala Asn Asn Leu Leu Ala Ala Ala Ile Asp Thr Arg Met Phe 445 450 455	1816
cac gaa cac acg caa aag aac gat gct act ttt tat aac aga ttg gtg His Glu His Thr Gln Lys Asn Asp Ala Thr Phe Tyr Asn Arg Leu Val 460 465 470 475	1864
ccc aag aag ggc ggc tcg cgg aag ttc acg aaa tca atg ctg aag agg	1912

Pro	Lys	Lys	Gly	Gly	Ser	Arg	Lys	Phe	Thr	Lys	Ser	Met	Leu	Lys	Arg	
				480					485					490		
ctc	cag	aag	ctg	gga	att	aac	aaa	aca	gac	cca	gac	agt	ttg	tct	cca	1960
Leu	Gln	Lys	Leu	Gly	Ile	Asn	Lys	Thr	Asp	Pro	Asp	Ser	Leu	Ser	Pro	
			495					500					505			
aaa	gaa	atc	aaa	cga	ttt	gca	cgc	cta	aac	ata	gac	acg	gat	acc	atc	2008
Lys	Glu	Ile	Lys	Arg	Phe	Ala	Arg	Leu	Asn	Ile	Asp	Thr	Asp	Thr	Ile	
		510					515					520				
act	atc	aaa	cgg	gtg	gtt	gat	gtc	aac	gat	aga	ctt	cta	cgg	cag	att	2056
Thr	Ile	Lys	Arg	Val	Val	Asp	Val	Asn	Asp	Arg	Leu	Leu	Arg	Gln	Ile	
		525				530					535					
acg	atc	ggc	caa	gct	cat	act	gag	aag	ggg	ttt	acc	aga	acc	acc	ggg	2104
Thr	Ile	Gly	Gln	Ala	His	Thr	Glu	Lys	Gly	Phe	Thr	Arg	Thr	Thr	Gly	
540					545					550					555	
ttt	gat	atc	act	gtt	gca	tct	gaa	ttg	atg	gcc	att	ttg	gca	ctt	tct	2152
Phe	Asp	Ile	Thr	Val	Ala	Ser	Glu	Leu	Met	Ala	Ile	Leu	Ala	Leu	Ser	
				560						565				570		
aag	gac	ttc	gct	gat	atg	aga	gaa	cgt	gta	ggg	cgc	atg	gtt	gtc	gct	2200
Lys	Asp	Phe	Ala	Asp	Met	Arg	Glu	Arg	Val	Gly	Arg	Met	Val	Val	Ala	
			575					580					585			
gcg	aat	cta	gac	ggc	gag	ccg	att	aca	gca	gaa	gat	gta	ggc	tgt	gct	2248
Ala	Asn	Leu	Asp	Gly	Glu	Pro	Ile	Thr	Ala	Glu	Asp	Val	Gly	Cys	Ala	
		590					595					600				
gga	gct	gtt	gct	gcc	ctt	ctg	aag	gac	gct	atc	aag	cca	aca	cta	atg	2296
Gly	Ala	Val	Ala	Ala	Leu	Leu	Lys	Asp	Ala	Ile	Lys	Pro	Thr	Leu	Met	
	605					610					615					
cag	tct	ttg	gaa	ggc	acc	cct	gtt	tta	gtc	cat	gca	ggg	ccg	ttc	gcc	2344
Gln	Ser	Leu	Glu	Gly	Thr	Pro	Val	Leu	Val	His	Ala	Gly	Pro	Phe	Ala	
	620				625					630					635	
aac	atc	tct	atc	gga	gcc	tcc	tca	gta	att	gct	gac	aga	att	gcc	ctg	2392
Asn	Ile	Ser	Ile	Gly	Ala	Ser	Ser	Val	Ile	Ala	Asp	Arg	Ile	Ala	Leu	
				640					645					650		
aag	ttg	gtg	ggc	act	tct	cca	cat	tct	ccg	gaa	ggc	acc	gac	gcc	ggc	2440
Lys	Leu	Val	Gly	Thr	Ser	Pro	His	Ser	Pro	Glu	Gly	Thr	Asp	Ala	Gly	
			655					660					665			
tat	gtg	gtt	acc	gag	gcc	gga	ttc	gac	ttt	aca	atg	ggc	gga	gaa	cga	2488
Tyr	Val	Val	Thr	Glu	Ala	Gly	Phe	Asp	Phe	Thr	Met	Gly	Gly	Glu	Arg	
		670					675					680				
ttc	ctg	aac	ata	aag	tgc	cgc	tcc	tca	ggc	ttg	aag	cca	aat	act	atc	2536
Phe	Leu	Asn	Ile	Lys	Cys	Arg	Ser	Ser	Gly	Leu	Lys	Pro	Asn	Thr	Ile	
	685					690					695					
gtc	atc	gtg	gcc	act	gtc	cgt	gcc	ctg	aag	ctg	cac	ggc	ggg	ggc	cct	2584
Val	Ile	Val	Ala	Thr	Val	Arg	Ala	Leu	Lys	Leu	His	Gly	Gly	Gly	Pro	
	700				705					710					715	
gag	gta	aag	ccg	ggc	cag	cca	ctt	act	ccc	cac	tat	act	gag	gaa	aac	2632
Glu	Val	Lys	Pro	Gly	Gln	Pro	Leu	Thr	Pro	His	Tyr	Thr	Glu	Glu	Asn	
				720						725					730	

ctc gag cta gtc aca aga ggt gcg gcc aac ctg cag aag caa ata aga Leu Glu Leu Val Thr Arg Gly Ala Ala Asn Leu Gln Lys Gln Ile Arg 735 740 745	2680
aat gcc aag ctt ttc ggg gtc cct gtc gta gtc gct gtg aac agg ttt Asn Ala Lys Leu Phe Gly Val Pro Val Val Val Ala Val Asn Arg Phe 750 755 760	2728
gag tca gac act gtc gca gaa att gag att att cag aag gcg gcg aag Glu Ser Asp Thr Val Ala Glu Ile Glu Ile Ile Gln Lys Ala Ala Lys 765 770 775	2776
gag gct ggt gca cat gac gcc atc gat gca aat cgc tgg gcc gaa ggc Glu Ala Gly Ala His Asp Ala Ile Asp Ala Asn Arg Trp Ala Glu Gly 780 785 790 795	2824
gga aaa gga gca gtc acg ctg gcc gaa gcc gtc atc aac tca tgc aaa Gly Lys Gly Ala Val Thr Leu Ala Glu Ala Val Ile Asn Ser Cys Lys 800 805 810	2872
caa ccc agt tct ttt aag ttc ctt tat gac gtc gat caa cct gtg aca Gln Pro Ser Ser Phe Lys Phe Leu Tyr Asp Val Asp Gln Pro Val Thr 815 820 825	2920
gaa aag ctc tcc acc att gtt cgg aaa atg tat ggc ggt tcc gcg ata Glu Lys Leu Ser Thr Ile Val Arg Lys Met Tyr Gly Gly Ser Ala Ile 830 835 840	2968
gat ctc tct cct acg gcg caa aag aag att gcc ttg tac acg gcc caa Asp Leu Ser Pro Thr Ala Gln Lys Lys Ile Ala Leu Tyr Thr Ala Gln 845 850 855	3016
ggt ttt gat aaa cta cct atc tgt atc gcg aaa act cag tac tcc ctg Gly Phe Asp Lys Leu Pro Ile Cys Ile Ala Lys Thr Gln Tyr Ser Leu 860 865 870 875	3064
tcc cat gat gcc act ttg aag aac gtc ccc tcc gat ttt gtg ttc ccc Ser His Asp Ala Thr Leu Lys Asn Val Pro Ser Asp Phe Val Phe Pro 880 885 890	3112
ata agg gat atc agg cct tcc att gga gcg ggc tac ctg tac gcg cta Ile Arg Asp Ile Arg Pro Ser Ile Gly Ala Gly Tyr Leu Tyr Ala Leu 895 900 905	3160
gcg gca gag att caa act atc ccc ggt ctc tcc act aac gca ggc tac Ala Ala Glu Ile Gln Thr Ile Pro Gly Leu Ser Thr Asn Ala Gly Tyr 910 915 920	3208
atg aat gtt gaa gtc agt gat gac ggc gag att gaa ggt ctt ttc Met Asn Val Glu Val Ser Asp Asp Gly Glu Ile Glu Gly Leu Phe 925 930 935	3253
taggtctacg attacgtatg cttagccacc cccgctcata tacataatta tggtgaacag	3313
gcatgctcag gctgcgttcc ctgtagtgca ccggttagtg cgcatgggcc ccgaagccgt	3373
tgctttttaga accggcatgc gacacatacc ggagacatta ctttagcgtc gcagtgcag	3433
ttatgtcccg agacgctggc atgcaatoga tgctgagcct agtgagtcag aaacggagtc	3493
agtgtcttca gacacacaag cgaaatcagg gtcagctttg actaaaccta aaaaccatag	3553

```

tgcattacgt ccgactgtat cttcacataa agttagggag gagaatcaaa gcaaggtagc 3613
ataacaaagt atggatttga tggtcataaa taaggacgac tgggtatatg tacagtccga 3673
gggcgaccac agcggcgatc acacttatTT tgagagcaca atcacctcat cgcttcgaga 3733
cacaaagcgg cacgcgctat tcccgcccat agtcagggag cagttcatcc aggggaaaga 3793
tatgtgccta ccaatctacc acgacgggat aaaagagcag gcagcgctcg agtacgacgg 3853
cacaggcgcg ctatgcggcc tttgacacgc tgtggcacca cgccgccaca cagcacgcaa 3913
taaacgggca acggtatatt gctgatatac ctcgatactt attattagaa aaaacctgcg 3973
tttgagggt tttatactac ataaagacag gatgaaaaac gtagatagat cagagtacgc 4033
tatggtttca gggtgataat ttatacaaaa tgcgcgccat tggaggaacc ctccagtgtc 4093
taggcatctt cctttagcaa agcctctctc ttttcagcga ttctttgagc tcttctctct 4153
ctggcagctc tgttcttcaa tcttctggcc tcagcctcct ccttcaaggg cttctcacgc 4213
tgagcatcag ccttggcctg gataatgtgt tcgaccaagg atctcttgtg cttgaaggtg 4273
ttacccttgg cagccttgta caaagagtgg tacaagtgtc tgtcgatctt gccggcgtct 4333
ctgtacttag ccaacaatct tctcaagaca cgcaatcttc tgatccagac gacctgggat 4393
ggcagtcctac cctccttggt accctttcct cttaccgtaa ccgggggtga cggggcgttt 4453
tcttggactc agccatggct c 4474

```

<210> 5

<211> 938

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 72

<400> 5

```

Tyr His Leu Ile Ser Gly Lys Ser Ile Ala Lys Thr Ile Arg Asp Asn
1           5           10           15

```

```

Ala Leu Arg Glu Val Lys Glu Ile Arg Gln Gln Val Pro Asp Phe Asn
20           25           30

```

```

Pro Lys Leu Lys Ile Leu Gln Val Gly Ser Arg Pro Asp Ser Ser Ala
35           40           45

```

Tyr Val Arg Met Lys Leu Lys Ala Ser Thr Glu Ser Gly Val Glu Cys
 50 55 60

Asp Val Glu Lys Leu Pro Glu Asp Ile Thr Gln Leu Glu Leu Leu Arg
 65 70 75 80

Lys Ile Glu Lys Ile Asn Ser Asp Ala Glu Val His Gly Ile Leu Ile
 85 90 95

Gln Leu Pro Leu Pro Lys His Leu Asp Glu Val Thr Val Thr Asn Ala
 100 105 110

Val Asp His Glu Lys Asp Val Asp Gly Phe His Arg Tyr Asn Val Gly
 115 120 125

Glu Leu Ala Lys Arg Gly Gly Lys Pro Tyr Phe Leu Pro Cys Thr Pro
 130 135 140

Asn Ala Cys Ile Gln Leu Leu Lys Glu Cys Gly Val Glu Ile Arg Gly
 145 150 155 160

Lys Gln Ala Val Val Ile Gly Arg Ser Asp Ile Val Gly Thr Pro Val
 165 170 175

Ala Thr Met Leu Lys Asn Leu Asp Ala Thr Val Thr Val Ala His Arg
 180 185 190

His Thr Thr Asn Leu Pro Gln Val Leu Ser Thr Ala Asp Ile Val Val
 195 200 205

Ala Ala Cys Gly Val Pro Asn Tyr Ile Lys Gly Glu Trp Leu Lys Asp
 210 215 220

Gly Ala Val Val Ile Asp Val Gly Ile Asn Tyr Val Gln Asp Ser Thr
 225 230 235 240

Lys Lys Ser Gly Gln Arg Leu Val Gly Asp Val Asp Phe Glu Ser Ala
 245 250 255

Lys Glu Lys Ala Ser His Ile Thr Pro Val Pro Gly Gly Val Gly Pro
 260 265 270

Met Thr Val Ala Met Leu Val Gln Asn Val Leu Thr Ala Ala Lys Arg
 275 280 285

Ser Leu Gln Lys Asp Ala Glu Val Pro Val Ile Thr Pro Leu Pro Val
 290 295 300

Asn Phe Glu Met Pro Val Pro Ser Asp Ile Asp Ile Ser Arg Lys Gln
 305 310 315 320

Asn Pro Lys His Ile Ser Gln Val Ala Ala Glu Leu Gly Ile Arg Asp
 325 330 335

His Glu Leu Glu Leu Phe Gly His Tyr Lys Ala Lys Ile Ser Pro Lys
 340 345 350

Leu Leu Gln Arg Leu His Ala Asn Gln Asn Ala Asn Tyr Val Leu Val
 355 360 365

Ala Gly Ile Thr Pro Thr Pro Leu Gly Glu Gly Lys Ser Thr Thr Thr
 370 375 380

Met Gly Leu Val Gln Ala Leu Ser Ala His Leu Gly Lys Arg Ala Ile
 385 390 395 400

Ala Asn Val Arg Gln Pro Ser Met Gly Pro Thr Phe Gly Val Lys Gly
 405 410 415

Gly Ala Ala Gly Gly Gly Tyr Ala Gln Val Ile Pro Met Asp Glu Phe
 420 425 430

Asn Leu His Leu Thr Gly Asp Ile His Ala Ile Ser Ala Ala Asn Asn
 435 440 445

Leu Leu Ala Ala Ala Ile Asp Thr Arg Met Phe His Glu His Thr Gln
 450 455 460

Lys Asn Asp Ala Thr Phe Tyr Asn Arg Leu Val Pro Lys Lys Gly Gly
 465 470 475 480

Ser Arg Lys Phe Thr Lys Ser Met Leu Lys Arg Leu Gln Lys Leu Gly
 485 490 495

Ile Asn Lys Thr Asp Pro Asp Ser Leu Ser Pro Lys Glu Ile Lys Arg
 500 505 510

Phe Ala Arg Leu Asn Ile Asp Thr Asp Thr Ile Thr Ile Lys Arg Val
 515 520 525

Val Asp Val Asn Asp Arg Leu Leu Arg Gln Ile Thr Ile Gly Gln Ala

530		535		540
His Thr Glu Lys Gly Phe Thr Arg Thr Thr Gly Phe Asp Ile Thr Val				
545		550		555
				560
Ala Ser Glu Leu Met Ala Ile Leu Ala Leu Ser Lys Asp Phe Ala Asp				
	565		570	575
Met Arg Glu Arg Val Gly Arg Met Val Val Ala Ala Asn Leu Asp Gly				
	580		585	590
Glu Pro Ile Thr Ala Glu Asp Val Gly Cys Ala Gly Ala Val Ala Ala				
	595		600	605
Leu Leu Lys Asp Ala Ile Lys Pro Thr Leu Met Gln Ser Leu Glu Gly				
	610		615	620
Thr Pro Val Leu Val His Ala Gly Pro Phe Ala Asn Ile Ser Ile Gly				
	625		630	635
				640
Ala Ser Ser Val Ile Ala Asp Arg Ile Ala Leu Lys Leu Val Gly Thr				
	645		650	655
Ser Pro His Ser Pro Glu Gly Thr Asp Ala Gly Tyr Val Val Thr Glu				
	660		665	670
Ala Gly Phe Asp Phe Thr Met Gly Gly Glu Arg Phe Leu Asn Ile Lys				
	675		680	685
Cys Arg Ser Ser Gly Leu Lys Pro Asn Thr Ile Val Ile Val Ala Thr				
	690		695	700
Val Arg Ala Leu Lys Leu His Gly Gly Gly Pro Glu Val Lys Pro Gly				
	705		710	715
				720
Gln Pro Leu Thr Pro His Tyr Thr Glu Glu Asn Leu Glu Leu Val Thr				
	725		730	735
Arg Gly Ala Ala Asn Leu Gln Lys Gln Ile Arg Asn Ala Lys Leu Phe				
	740		745	750
Gly Val Pro Val Val Val Ala Val Asn Arg Phe Glu Ser Asp Thr Val				
	755		760	765
Ala Glu Ile Glu Ile Ile Gln Lys Ala Ala Lys Glu Ala Gly Ala His				
	770		775	780

Asp Ala Ile Asp Ala Asn Arg Trp Ala Glu Gly Gly Lys Gly Ala Val
785 790 795 800

Thr Leu Ala Glu Ala Val Ile Asn Ser Cys Lys Gln Pro Ser Ser Phe
805 810 815

Lys Phe Leu Tyr Asp Val Asp Gln Pro Val Thr Glu Lys Leu Ser Thr
820 825 830

Ile Val Arg Lys Met Tyr Gly Gly Ser Ala Ile Asp Leu Ser Pro Thr
835 840 845

Ala Gln Lys Lys Ile Ala Leu Tyr Thr Ala Gln Gly Phe Asp Lys Leu
850 855 860

Pro Ile Cys Ile Ala Lys Thr Gln Tyr Ser Leu Ser His Asp Ala Thr
865 870 875 880

Leu Lys Asn Val Pro Ser Asp Phe Val Phe Pro Ile Arg Asp Ile Arg
885 890 895

Pro Ser Ile Gly Ala Gly Tyr Leu Tyr Ala Leu Ala Ala Glu Ile Gln
900 905 910

Thr Ile Pro Gly Leu Ser Thr Asn Ala Gly Tyr Met Asn Val Glu Val
915 920 925

Ser Asp Asp Gly Glu Ile Glu Gly Leu Phe
930 935

<210> 6

<211> 918

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 81

<400> 6

gatcgattcg ctttcgtgag acacttggcg taataccaca cgcgagccgc aatgggcccgc

60

```

cttaacgagg tagtgggcat atagtttctc cttctgagtt agcctgttca aagaattttc 120
gagcgctgag cataataacg ggagcattcc gttccgcaa gtataactga ctgagctggc 180
tgttcgtagt agccatcggt cttagaatgc taaatgtatt acgcgaatgt atcgatactc 240
gcttaaatat cgactgtagc atataaaaac gaacgaagca gttggtcctg gtccttattt 300
atgtacaagc attacgttca aagttagtga ccaatgttgc aaaaagcctt tttggaggat 360
ttcatagcgc cgtttcctcg gggggaggaa gactacttta atttaggcaa gtaaagtctg 420
cgccagcctc aactgagtca cgccatcaga tgacgacgga tggatactca tctagccgca 480
agtattagtc gctgatttac ggcgcgttct attgatctga ttggctactt cggaatgtaa 540
atagaattta caaagtcagg gattcacctt tgtccttctt ggattgaccg tactttttca 600
tgogaattgc ttgaatagcg atgaaaccag tggtatcagt cggctggaat ccggtcaatt 660
cgtccatgga ggactccgtc gcatcataca agttctcagt ttccgaggac ctgccaaca 720
caatgacatt acccttgtaa agtctcaagc gaacctggcc atttaccgta atttgggatg 780
gcttaatcat gctacggata tactctgcat tccggagaaa agtaggagcc attgtagagg 840
aggcgagagt aagtcggctg taacaaattg gtcacgcaaa tgccgaactt ctttgtccag 900
ggtgagaccc tcgagatc 918

```

<210> 7

<211> 20

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 81

<400> 7

```

Asp Leu Glu Gly Leu Thr Leu Asp Lys Glu Val Arg His Leu Arg Asp
1           5           10          15

```

```

Gln Phe Val Thr
          20

```

<210> 8

<211> 104

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 81

<400> 8

Tyr Ser Arg Leu Thr Leu Ala Ser Ser Thr Met Ala Pro Thr Phe Leu
1 5 10 15

Arg Asn Ala Glu Tyr Ile Arg Ser Met Ile Lys Pro Ser Gln Ile Thr
20 25 30

Val Asn Gly Gln Val Arg Leu Arg Leu Tyr Lys Gly Asn Val Ile Val
35 40 45

Leu Gly Arg Ser Ser Glu Thr Glu Asn Leu Tyr Asp Ala Thr Glu Ser
50 55 60

Ser Met Asp Glu Leu Thr Gly Phe Gln Pro Thr Asp Thr Thr Gly Phe
65 70 75 80

Ile Ala Ile Gln Ala Ile Arg Met Lys Lys Tyr Gly Gln Ser Lys Lys
85 90 95

Asp Lys Gly Glu Ser Leu Thr Leu
100

<210> 9

<211> 2898

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (716) .. (1957)

<223>

<220>

<221> CDS

<222> (2246) .. (2839)

<223>

<220>

<221> misc_feature

<223> Oligo 81

<400> 9

```

gtgcagtaact tcatctgcgt gagaaggggt tcgggcgcac atggttggtta aatgaaatgc      60
cgtttgttca gaggcgtatc ttctgtagtc aggaactccc gcccgccaa cgaatgggaa      120
acgacaaatt gcgggaacgc gggccttgcc gcacagagtc ttgtataaac aagatcgccc      180
tcctggtaag atataactgg ggcactcatc gcacacttta cccacgcggc ttgtactgat      240
ctctctgagc ttgccgacgc cctgaacttt ctgcctgttg caagccatca cgtgacaaaa      300
tattacatat cacgtgttac aagcagaacc agtctgagta cggttatgca ttggtccata      360
atctgcgcgt gagggtcgcg ctgcccgcta ctttgtactc aatgtccac ttgtgcactt      420
ttctttcgtg tcctgcggca ggagtcaccc gccccaatag ttatgctgtc ttagccgcct      480
agctacgttg gcgtacacac gtagcgagca ccaacagttt ggtagatcaa gccttaccta      540
gctctgtgac tcaaccattc agtcatagtt ctgtttttct gctagtggtc tggtagtgta      600
ctcttgcgag aatggctatg agtcactttt tttgattcat gtgataagct tgttcggagt      660
atataaacat cccttacgag tttcgtttgg aagatctctc acatccggaa gcaag atg      718
                                         Met
                                         1
gca cgc gga aag gtt tgc ttg gca tac tca gga ggc ctg gac act tcg      766
Ala Arg Gly Lys Val Cys Leu Ala Tyr Ser Gly Gly Leu Asp Thr Ser
          5              10              15

gtg atc ctc gcc tgg tta ctg gag cag ggt tat gaa gtt gta gcg ttt      814
Val Ile Leu Ala Trp Leu Leu Glu Gln Gly Tyr Glu Val Val Ala Phe
          20              25              30

atg gct aat gtt ggg caa gag gaa gat ttc cag gct gcg gaa gag aag      862
Met Ala Asn Val Gly Gln Glu Glu Asp Phe Gln Ala Ala Glu Glu Lys
          35              40              45

gca ttg aag ata gga gcg tcc aag ttt gtg tgc gtt gac tgc cgt gag      910

```

Ala 50	Leu	Lys	Ile	Gly	Ala 55	Ser	Lys	Phe	Val	Cys 60	Val	Asp	Cys	Arg	Glu 65	
cag	ttc	gtg	acg	gat	gtc	ttg	ttt	cct	gcg	gtg	cag	gtg	aac	gcg	gtg	958
Gln	Phe	Val	Thr	Asp	Val	Leu	Phe	Pro	Ala	Val	Gln	Val	Asn	Ala	Val	
			70					75						80		
tac	gaa	gat	gtt	tat	cta	ctg	ggg	acg	tcg	ctt	gca	cgg	ccc	gtg	att	1006
Tyr	Glu	Asp	Val	Tyr	Leu	Leu	Gly	Thr	Ser	Leu	Ala	Arg	Pro	Val	Ile	
			85					90					95			
gcc	aag	gca	cag	atc	gac	atc	gcg	gag	cag	gag	cat	tgt	ttt	gcg	gtc	1054
Ala	Lys	Ala	Gln	Ile	Asp	Ile	Ala	Glu	Gln	Glu	His	Cys	Phe	Ala	Val	
		100					105					110				
gct	cac	gga	tgc	act	ggg	aag	ggc	aac	gat	cag	atc	cgg	ttt	gag	ctt	1102
Ala	His	Gly	Cys	Thr	Gly	Lys	Gly	Asn	Asp	Gln	Ile	Arg	Phe	Glu	Leu	
	115					120					125					
gcg	ttc	tac	gcg	ctg	aag	cca	gac	gtc	cag	gtt	atc	gcg	cca	tgg	cgg	1150
Ala	Phe	Tyr	Ala	Leu	Lys	Pro	Asp	Val	Gln	Val	Ile	Ala	Pro	Trp	Arg	
130					135					140					145	
gac	cct	gcg	ttt	ttc	gac	cgg	ttt	gca	ggt	aga	aat	gac	ttg	cta	gag	1198
Asp	Pro	Ala	Phe	Phe	Asp	Arg	Phe	Ala	Gly	Arg	Asn	Asp	Leu	Leu	Glu	
				150					155					160		
tgt	gcg	gca	gag	aaa	cag	ata	cct	gtt	gcg	cag	acc	aag	gcc	aaa	ccg	1246
Cys	Ala	Ala	Glu	Lys	Gln	Ile	Pro	Val	Ala	Gln	Thr	Lys	Ala	Lys	Pro	
			165					170					175			
tgg	tct	acc	gac	gag	aac	gag	gcg	cat	ata	tcg	tac	gag	gct	gga	att	1294
Trp	Ser	Thr	Asp	Glu	Asn	Glu	Ala	His	Ile	Ser	Tyr	Glu	Ala	Gly	Ile	
		180					185					190				
ttg	gag	gat	ccc	gat	gtc	aca	cca	ccg	gca	gac	atg	tgg	aag	ctc	acg	1342
Leu	Glu	Asp	Pro	Asp	Val	Thr	Pro	Pro	Ala	Asp	Met	Trp	Lys	Leu	Thr	
	195					200					205					
aca	gac	ccg	ctt	cga	gca	cca	agt	gag	cct	gag	gac	gtg	acg	ttg	gtg	1390
Thr	Asp	Pro	Leu	Arg	Ala	Pro	Ser	Glu	Pro	Glu	Asp	Val	Thr	Leu	Val	
210					215					220					225	
ttc	aag	gag	ggt	gtc	ccc	act	cag	atg	att	tac	aag	gat	gga	gag	gcg	1438
Phe	Lys	Glu	Gly	Val	Pro	Thr	Gln	Met	Ile	Tyr	Lys	Asp	Gly	Glu	Ala	
				230					235					240		
agc	aaa	ccg	ttg	gat	att	ttt	ctt	gcg	gct	tcc	aaa	gtg	gcg	cgt	aga	1486
Ser	Lys	Pro	Leu	Asp	Ile	Phe	Leu	Ala	Ala	Ser	Lys	Val	Ala	Arg	Arg	
			245					250					255			
aac	ggc	gtt	ggc	aga	atc	gac	atc	gtg	gaa	aac	cgt	tac	atc	aac	ctg	1534
Asn	Gly	Val	Gly	Arg	Ile	Asp	Ile	Val	Glu	Asn	Arg	Tyr	Ile	Asn	Leu	
		260					265					270				
aag	tcg	cgt	ggc	tgc	tac	gaa	cag	gca	cca	ctg	aca	ctc	ctg	cgc	cgg	1582
Lys	Ser	Arg	Gly	Cys	Tyr	Glu	Gln	Ala	Pro	Leu	Thr	Leu	Leu	Arg	Arg	
	275					280					285					
gcc	cac	gtg	tat	ctc	gag	ggt	ctc	acc	ctg	gac	aaa	gag	gtt	cgg	cat	1630
Ala	His	Val	Tyr	Leu	Glu	Gly	Leu	Thr	Leu	Asp	Lys	Glu	Val	Arg	His	
290					295					300					305	

ttg cgt gac caa ttt gtt acg ccg act tac tct cgc ctc ctc tac aat Leu Arg Asp Gln Phe Val Thr Pro Thr Tyr Ser Arg Leu Leu Tyr Asn 310 315 320	1678
ggc tcc tac ttt tct ccg gaa tgc gag tat atc cgt agc atg att aag Gly Ser Tyr Phe Ser Pro Glu Cys Glu Tyr Ile Arg Ser Met Ile Lys 325 330 335	1726
cca tcc caa att acg gta aat ggc cag gtt cgc ttg aga ctt tac aag Pro Ser Gln Ile Thr Val Asn Gly Gln Val Arg Leu Arg Leu Tyr Lys 340 345 350	1774
ggt aat gtc att gtg ttg ggc agg tcc tcg gaa act gag aac ttg tat Gly Asn Val Ile Val Leu Gly Arg Ser Ser Glu Thr Glu Asn Leu Tyr 355 360 365	1822
gat gcg acg gag tcc tcc atg gac gaa ttg acc gga ttc cag ccg act Asp Ala Thr Glu Ser Ser Met Asp Glu Leu Thr Gly Phe Gln Pro Thr 370 375 380 385	1870
gat acc act ggt ttc atc gct att caa gca att cgc atg aaa aag tac Asp Thr Thr Gly Phe Ile Ala Ile Gln Ala Ile Arg Met Lys Lys Tyr 390 395 400	1918
ggt caa tcc aag aag gac aaa ggt gaa tcc ctg act ttg taaattctat Gly Gln Ser Lys Lys Asp Lys Gly Glu Ser Leu Thr Leu 405 410	1967
ttacattccg aagtagccaa tcagatcaat agaacggcgc gtaaatacagc gactaatact	2027
tgccggctaga tgagtatcca tccgtcgtca tctgatggcg tgactcagtt gaggctggcg	2087
cagactttac ttgcctaaat taaagtagtc ttctccccc cgaggaaacg gcgctatgaa	2147
atcctccaaa aaggcttttt gcaacattgg tcaactaactt tgaacgtaat gcttgtacat	2207
aaataaggac caggaccaac tgcttcgttc gtttttat atg cta cag tcg ata ttt Met Leu Gln Ser Ile Phe 415 420	2263
aag cga gta tcg ata cat tcg cgt aat aca ttt agc att cta aga acg Lys Arg Val Ser Ile His Ser Arg Asn Thr Phe Ser Ile Leu Arg Thr 425 430 435	2311
atg gct act acg aac aac cag ctc agt cag tta tac ttg gcg gaa ccg Met Ala Thr Thr Asn Asn Gln Leu Ser Gln Leu Tyr Leu Ala Glu Arg 440 445 450	2359
aat gct ccc gtt att atg ctc agc gct cga aaa ttc ttt gaa cag cta Asn Ala Pro Val Ile Met Leu Ser Ala Arg Lys Phe Phe Glu Gln Leu 455 460 465	2407
act cag aag gag aaa cta tat gcc cac tac ctc gtt aag gcg gcc cat Thr Gln Lys Glu Lys Leu Tyr Ala His Tyr Leu Val Lys Ala Ala His 470 475 480	2455
tcg ggc tcg cgt gtg gta tta cgc caa gtg tct cac gaa agc gaa tcg Cys Gly Ser Arg Val Val Leu Arg Gln Val Ser His Glu Ser Glu Ser 485 490 495 500	2503
atc ttt gat atg tta cta agt atc cat aaa gtg atg gcc ggg cgc tac	2551

```

Ile Phe Asp Met Leu Leu Ser Ile His Lys Val Met Ala Gly Arg Tyr
      505                      510                      515

cat gaa tca gaa gag aat act ctc ttc tta gag tat act tct caa ttc      2599
His Glu Ser Glu Glu Asn Thr Leu Phe Leu Glu Tyr Thr Ser Gln Phe
      520                      525                      530

ctt agc aac ttg ggg aac ttc aat tca ttt ggt gac act aag ttt att      2647
Leu Ser Asn Leu Gly Asn Phe Asn Ser Phe Gly Asp Thr Lys Phe Ile
      535                      540                      545

ccg cgc tgt acc caa ggg cac ttc ttt tcc ctt ttg aaa caa gcc ggc      2695
Pro Arg Cys Thr Gln Gly His Phe Phe Ser Leu Leu Lys Gln Ala Gly
      550                      555                      560

ctc gac aag aaa gcg aag cct gcc act aat gga tgt cca ttt aaa acc      2743
Leu Asp Lys Lys Ala Lys Pro Ala Thr Asn Gly Cys Pro Phe Lys Thr
565                      570                      575                      580

tat aaa gaa tta gtg gag att ggc gtc tat gcc gtt gat gcc caa atc      2791
Tyr Lys Glu Leu Val Glu Ile Gly Val Tyr Ala Val Asp Ala Gln Ile
      585                      590                      595

tgc aat gct aag ttt ccc gga tct tgg tgc tac gac ggc tta tta ctt      2839
Cys Asn Ala Lys Phe Pro Gly Ser Trp Ser Tyr Asp Gly Leu Leu Leu
      600                      605                      610

taactgcggtt agtaaagccg~acatgcagtt ggttgcagaa ggaggtcttt gccgcctc      2898

```

<210> 10

<211> 414

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 81

<400> 10

```

Met Ala Arg Gly Lys Val Cys Leu Ala Tyr Ser Gly Gly Leu Asp Thr
1                      5                      10                      15

```

```

Ser Val Ile Leu Ala Trp Leu Leu Glu Gln Gly Tyr Glu Val Val Ala
      20                      25                      30

```

```

Phe Met Ala Asn Val Gly Gln Glu Glu Asp Phe Gln Ala Ala Glu Glu
      35                      40                      45

```

```

Lys Ala Leu Lys Ile Gly Ala Ser Lys Phe Val Cys Val Asp Cys Arg
      50                      55                      60

```

Glu Gln Phe Val Thr Asp Val Leu Phe Pro Ala Val Gln Val Asn Ala
65 70 75 80

Val Tyr Glu Asp Val Tyr Leu Leu Gly Thr Ser Leu Ala Arg Pro Val
85 90 95

Ile Ala Lys Ala Gln Ile Asp Ile Ala Glu Gln Glu His Cys Phe Ala
100 105 110

Val Ala His Gly Cys Thr Gly Lys Gly Asn Asp Gln Ile Arg Phe Glu
115 120 125

Leu Ala Phe Tyr Ala Leu Lys Pro Asp Val Gln Val Ile Ala Pro Trp
130 135 140

Arg Asp Pro Ala Phe Phe Asp Arg Phe Ala Gly Arg Asn Asp Leu Leu
145 150 155 160

Glu Cys Ala Ala Glu Lys Gln Ile Pro Val Ala Gln Thr Lys Ala Lys
165 170 175

Pro Trp Ser Thr Asp Glu Asn Glu Ala His Ile Ser Tyr Glu Ala Gly
180 185 190

Ile Leu Glu Asp Pro Asp Val Thr Pro Pro Ala Asp Met Trp Lys Leu
195 200 205

Thr Thr Asp Pro Leu Arg Ala Pro Ser Glu Pro Glu Asp Val Thr Leu
210 215 220

Val Phe Lys Glu Gly Val Pro Thr Gln Met Ile Tyr Lys Asp Gly Glu
225 230 235 240

Ala Ser Lys Pro Leu Asp Ile Phe Leu Ala Ala Ser Lys Val Ala Arg
245 250 255

Arg Asn Gly Val Gly Arg Ile Asp Ile Val Glu Asn Arg Tyr Ile Asn
260 265 270

Leu Lys Ser Arg Gly Cys Tyr Glu Gln Ala Pro Leu Thr Leu Leu Arg
275 280 285

Arg Ala His Val Tyr Leu Glu Gly Leu Thr Leu Asp Lys Glu Val Arg
290 295 300

His Leu Arg Asp Gln Phe Val Thr Pro Thr Tyr Ser Arg Leu Leu Tyr
305 310 315 320

Asn Gly Ser Tyr Phe Ser Pro Glu Cys Glu Tyr Ile Arg Ser Met Ile
325 330 335

Lys Pro Ser Gln Ile Thr Val Asn Gly Gln Val Arg Leu Arg Leu Tyr
340 345 350

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

Tyr Asp Ala Thr Glu Ser Ser Met Asp Glu Leu Thr Gly Phe Gln Pro
370 375 380

Thr Asp Thr Thr Gly Phe Ile Ala Ile Gln Ala Ile Arg Met Lys Lys
385 390 395 400

Tyr Gly Gln Ser Lys Lys Asp Lys Gly Glu Ser Leu Thr Leu
405 410

<210> 11

<211> 198

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 81

<400> 11

Met Leu Gln Ser Ile Phe Lys Arg Val Ser Ile His Ser Arg Asn Thr
1 5 10 15

Phe Ser Ile Leu Arg Thr Met Ala Thr Thr Asn Asn Gln Leu Ser Gln
20 25 30

Leu Tyr Leu Ala Glu Arg Asn Ala Pro Val Ile Met Leu Ser Ala Arg
35 40 45

Lys Phe Phe Glu Gln Leu Thr Gln Lys Glu Lys Leu Tyr Ala His Tyr
50 55 60

Leu Val Lys Ala Ala His Cys Gly Ser Arg Val Val Leu Arg Gln Val
65 70 75 80

Ser His Glu Ser Glu Ser Ile Phe Asp Met Leu Leu Ser Ile His Lys
85 90 95

Val Met Ala Gly Arg Tyr His Glu Ser Glu Glu Asn Thr Leu Phe Leu
100 105 110

Glu Tyr Thr Ser Gln Phe Leu Ser Asn Leu Gly Asn Phe Asn Ser Phe
115 120 125

Gly Asp Thr Lys Phe Ile Pro Arg Cys Thr Gln Gly His Phe Phe Ser
130 135 140

Leu Leu Lys Gln Ala Gly Leu Asp Lys Lys Ala Lys Pro Ala Thr Asn
145 150 155 160

Gly Cys Pro Phe Lys Thr Tyr Lys Glu Leu Val Glu Ile Gly Val Tyr
165 170 175

Ala Val Asp Ala Gln Ile Cys Asn Ala Lys Phe Pro Gly Ser Trp Ser
180 185 190

Tyr Asp Gly Leu Leu Leu
195

<210> 12

<211> 139

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 86

<400> 12

gatccggtca gtagccacaa acagcaacgt ctcctgatca acctcgtaga tatcccgcac 60

cttaccacgc gccactaacg gcaggatgcc gccagctgt gtctgcacta tgctcatttt 120

tgttcgccgc cccagatc 139

<210> 13

<211> 39

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 86

<400> 13

Met Ser Ile Val Gln Thr Gln Leu Gly Gly Ile Leu Pro Leu Val Ala
1 5 10 15

Arg Gly Lys Val Arg Asp Ile Tyr Glu Val Asp Gln Glu Thr Leu Leu
20 25 30

Phe Val Ala Thr Asp Arg Ile
35

<210> 14

<211> 1671

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (587) .. (1492)

<223>

<220>

<221> misc_feature

<223> Oligo 86

<400> 14

taacccttgc ccggttcgct gcgctgcttc gcggtgcttc tggttttcgta accgtttcga 60

tctacgaaat gtgtgcaaca agccgaaaaa gcaaaccacc cgctccactg actatctgtt	120
tatgcttaga cgccgcaaac ttgtttctct ttgggcctac tctttgattc ctggctcaga	180
gccttcccta agtaaggaat ttcaactaot tgtatatcaa agccttgcta caattgcctt	240
ctacggcctc aattaccacg aatgcactaa gcagctcagc tgttccccctt gttatgaaac	300
gaaatcatca acagctaaca aatgcctagc tttactcctg tgacaagctg ccttcacgtt	360
tctttttcaa ccgttgacaa tgacgattaa gtaacaatgc gtctcgagga atttgctcga	420
ggtcgggtgg ggcagccgcg aaatgtagaa tgcacggcag tatgactcca ggcagcgca	480
gcttgactct ggtctatggc caatttaaca gccccatata gacaagatat atatagagag	540
atagcagata gtcccatctg gaacgatctg ggggcggcga acaaaa atg agc ata	595
Met Ser Ile	
1	
gtg cag aca cag ctg ggc ggc atc ctg ccg tta gtg gcg cgt ggt aag	643
Val Gln Thr Gln Leu Gly Gly Ile Leu Pro Leu Val Ala Arg Gly Lys	
5 10 15	
gtg cgg gat atc tac gag gtt gat cag gag acg ttg ctg ttt gtg gct	691
Val Arg Asp Ile Tyr Glu Val Asp Gln Glu Thr Leu Leu Phe Val Ala	
20 25 30 35	
act gac cgg atc tcg gcc tat gat gtg atc atg gcg aat gct atc ccc	739
Thr Asp Arg Ile Ser Ala Tyr Asp Val Ile Met Ala Asn Ala Ile Pro	
40 45 50	
gac aag ggt gtg ctg ttg acg cgg ctg tct gag ttc tgg ttt aaa ttc	787
Asp Lys Gly Val Leu Leu Thr Arg Leu Ser Glu Phe Trp Phe Lys Phe	
55 60 65	
ttg gag aaa gac gtc cgc aac cac ctc cgg gcg ccg tgt gag ggg ctt	835
Leu Glu Lys Asp Val Arg Asn His Leu Arg Ala Pro Cys Glu Gly Leu	
70 75 80	
ttt ggt ggt cta cct gat gcg cta acg tca ccc gag tac cgc tca cag	883
Phe Gly Gly Leu Pro Asp Ala Leu Thr Ser Pro Glu Tyr Arg Ser Gln	
85 90 95	
cta gag ggg cgg agc ctg ctc gtg caa cgg tac cgg ttg ata cca ctt	931
Leu Glu Gly Arg Ser Leu Leu Val Gln Arg Tyr Arg Leu Ile Pro Leu	
100 105 110 115	
gag gtc atc gtg cgc ggt tac att acg ggc tct gcc tgg aag gag tac	979
Glu Val Ile Val Arg Gly Tyr Ile Thr Gly Ser Ala Trp Lys Glu Tyr	
120 125 130	
cag gag cgc ggc aca gtg cat ggc ctg gct cta cct gcc ggg ctg cgt	1027
Gln Glu Arg Gly Thr Val His Gly Leu Ala Leu Pro Ala Gly Leu Arg	
135 140 145	
gag tct gag gag tta ccc gag ccg atg ttc act ccg tct acc aaa gcg	1075
Glu Ser Glu Glu Leu Pro Glu Pro Met Phe Thr Pro Ser Thr Lys Ala	
150 155 160	
gag cag ggt gcg cac gat gag aat ata tca cca gca cag gct gca gac	1123

Glu Gln Gly Ala His Asp Glu Asn Ile Ser Pro Ala Gln Ala Ala Asp
 165 170 175
 cta att ggg cag gag ctt tgt gac cgc gtg gct gcc ttg gca att cgc 1171
 Leu Ile Gly Gln Glu Leu Cys Asp Arg Val Ala Ala Leu Ala Ile Arg
 180 185 190 195
 ttg tat tcg aag tgc aag aag tat gcc agg gag cgc ggc att att att 1219
 Leu Tyr Ser Lys Cys Lys Lys Tyr Ala Arg Glu Arg Gly Ile Ile Ile
 200 205 210
 gca gac acg aag ttt gag ttt ggc att gat gac aag tca ggc gag att 1267
 Ala Asp Thr Lys Phe Glu Phe Gly Ile Asp Asp Lys Ser Gly Glu Ile
 215 220 225
 gtt ctt gtc gat gag gtc ttg acc cca gat tct tct cgt ttc tgg aat 1315
 Val Leu Val Asp Glu Val Leu Thr Pro Asp Ser Ser Arg Phe Trp Asn
 230 235 240
 gct gcg aac tac gaa gtt ggt aag cct cag gat tca tat gat aaa cag 1363
 Ala Ala Asn Tyr Glu Val Gly Lys Pro Gln Asp Ser Tyr Asp Lys Gln
 245 250 255
 ttt ttg cgc aat tgg ctc act agc aac aag ttg aat ggt gtc gat ggg 1411
 Phe Leu Arg Asn Trp Leu Thr Ser Asn Lys Leu Asn Gly Val Asp Gly
 260 265 270 275
 gtc agc atg ccc gag gag atc gtg acg cgc acg aag gag aaa tat gtt 1459
 Val Ser Met Pro Glu Glu Ile Val Thr Arg Thr Lys Glu Lys Tyr Val
 280 285 290
 gag gcc tac gaa gca ctt aca ggc gcg aaa tgg taaaatgatg tacttttaaa 1512
 Glu Ala Tyr Glu Ala Leu Thr Gly Ala Lys Trp
 295 300
 atgttaaata attttaagt ataaatggag atgtgaaatt tatgacctcc atgggttagt 1572
 tgggggtgcgc acatgggtacc ttgcatggat atctgggtacg tcctgaatac gtttgccact 1632
 caccagctcc cggggccgct tttctactgc caactcaaa 1671

 <210> 15
 <211> 302
 <212> PRT
 <213> Ashbya gossypii

 <220>
 <221> misc_feature
 <223> Oligo 86
 <400> 15
 Met Ser Ile Val Gln Thr Gln Leu Gly Gly Ile Leu Pro Leu Val Ala
 1 5 10 15

Arg Gly Lys Val Arg Asp Ile Tyr Glu Val Asp Gln Glu Thr Leu Leu
20 25 30

Phe Val Ala Thr Asp Arg Ile Ser Ala Tyr Asp Val Ile Met Ala Asn
35 40 45

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

Phe Lys Phe Leu Glu Lys Asp Val Arg Asn His Leu Arg Ala Pro Cys
65 70 75 80

Glu Gly Leu Phe Gly Gly Leu Pro Asp Ala Leu Thr Ser Pro Glu Tyr
85 90 95

Arg Ser Gln Leu Glu Gly Arg Ser Leu Leu Val Gln Arg Tyr Arg Leu
100 105 110

Ile Pro Leu Glu Val Ile Val Arg Gly Tyr Ile Thr Gly Ser Ala Trp
115 120 125

Lys Glu Tyr Gln Glu Arg Gly Thr Val His Gly Leu Ala Leu Pro Ala
130 135 140

Gly Leu Arg Glu Ser Glu Glu Leu Pro Glu Pro Met Phe Thr Pro Ser
145 150 155 160

Thr Lys Ala Glu Gln Gly Ala His Asp Glu Asn Ile Ser Pro Ala Gln
165 170 175

Ala Ala Asp Leu Ile Gly Gln Glu Leu Cys Asp Arg Val Ala Ala Leu
180 185 190

Ala Ile Arg Leu Tyr Ser Lys Cys Lys Lys Tyr Ala Arg Glu Arg Gly
195 200 205

Ile Ile Ile Ala Asp Thr Lys Phe Glu Phe Gly Ile Asp Asp Lys Ser
210 215 220

Gly Glu Ile Val Leu Val Asp Glu Val Leu Thr Pro Asp Ser Ser Arg
225 230 235 240

Phe Trp Asn Ala Ala Asn Tyr Glu Val Gly Lys Pro Gln Asp Ser Tyr
245 250 255

Asp Lys Gln Phe Leu Arg Asn Trp Leu Thr Ser Asn Lys Leu Asn Gly
 260 265 270

Val Asp Gly Val Ser Met Pro Glu Glu Ile Val Thr Arg Thr Lys Glu
 275 280 285

Lys Tyr Val Glu Ala Tyr Glu Ala Leu Thr Gly Ala Lys Trp
 290 295 300

<210> 16

<211> 1325

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 162

<400> 16
 gatctcgaac tggcggatga tgcagtccaa ggactacgct acatggacaa ggccgggggca 60
 cagcacatat tgcacacgaa gaatgtagtt ttctgtactg gcgggttttg gtcgtccaag 120
 gagttgttgg ggaagtatgc ccccatctc cgggaccttc ctacaacaaa cgcgagggt 180
 actacgggag atggacagga cctgatgctc cgtgttggtg ggcagctgat tgatatggag 240
 catatccagg tacacccac gagttttgtg gatcctaaag ataaagacag cggtagcaaa 300
 ttctctggcgg ctgaggcact gcgcggcaat ggagggtgtgc ttttaaacc cagcactggc 360
 cggcgcttta cgaacgaatt ggcgacacgt gatgtactta cggctagcat tcaagagcac 420
 tgcaaagagc atgtggctta ccttgttctg agccaagggg catatgaaac tttgcaaagc 480
 tttgtcaact tctacatctc taaagggctg atgcgagcgt ccaccgttg cgagctttcc 540
 gacgagatcg gagttccgaa acaacagctg gcgtcggagc tagaggcgta ctccgccgca 600
 gaggtggatc aattccagcg tcggcacatt atcaatgatt ttggcccgag tgttgatggg 660
 tcgacagcaa tatatgttgg cttgatcaca cccgcagtc attttaccat gggcgggtgtt 720
 cgaatcaatt ccagagcgga agtgctgagc agcgcaggga cgacctataa agggctttac 780
 gccgctggag aggtatccgg cggcgtgcac ggtagaaaca ggctgggtgg gtctagtctc 840
 ctggagtgcg tcgttttcgg gaggcaggca gcaacatcga ttacagaaag gttgcaggcg 900

```

tgaactctag caaccctggc tcagtaacag acagtttttg cctgaaatgt gtcattttaa 960
taacagcaga tagtaggtag catacaaaat taattagaga cgtcaattaa tactgcaatt 1020
aggaaactag ggatatcaac tataattatt aaaaacgtcg agtaaaaaga attaattgga 1080
cggtccttga aatagagccg aagaagctga agaagaacag taaatagtga atacgatgag 1140
aaaggccttc aggtgccgct tttgagtggg tagacataaa gcatttttggg ttcttagatc 1200
aatcgaaagg aatgtgctgc cttggatatt ttgatgctct ttttaggtac aggtggactg 1260
ccaaaaaatg gcttgttcgc tagcctaatg ggtattattg cgtgaccaat ctgccgtagt 1320
cagtt 1325

```

<210> 17

<211> 294

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 162

<400> 17

```

Asp Leu Glu Leu Ala Asp Asp Ala Val Gln Gly Leu Arg Tyr Met Asp
1           5           10           15

```

```

Lys Ala Gly Ala Gln His Ile Leu His Thr Lys Asn Val Val Phe Cys
20           25           30

```

```

Thr Gly Gly Phe Gly Ser Ser Lys Glu Leu Leu Gly Lys Tyr Ala Pro
35           40           45

```

```

His Leu Arg Asp Leu Pro Thr Thr Asn Ala Gln Gly Thr Thr Gly Asp
50           55           60

```

```

Gly Gln Asp Leu Met Leu Arg Val Gly Gly Gln Leu Ile Asp Met Glu
65           70           75           80

```

```

His Ile Gln Val His Pro Thr Ser Phe Val Asp Pro Lys Asp Lys Asp
85           90           95

```

```

Ser Gly Ser Lys Phe Leu Ala Ala Glu Ala Leu Arg Gly Asn Gly Gly

```

100	105	110
Val Leu Leu Asn Pro Ser Thr Gly Arg Arg Phe Thr Asn Glu Leu Ala 115 120 125		
Thr Arg Asp Val Leu Thr Ala Ser Ile Gln Glu His Cys Lys Glu His 130 135 140		
Val Ala Tyr Leu Val Leu Ser Gln Gly Ala Tyr Glu Thr Leu Gln Ser 145 150 155 160		
Phe Val Asn Phe Tyr Ile Ser Lys Gly Leu Met Arg Ala Ser Thr Val 165 170 175		
Gly Glu Leu Ser Asp Glu Ile Gly Val Pro Lys Gln Gln Leu Ala Ser 180 185 190		
Glu Leu Glu Ala Tyr Ser Ala Ala Glu Val Asp Gln Phe Gln Arg Arg 195 200 205		
His Ile Ile Asn Asp Phe Gly Pro Ser Val Asp Gly Ser Thr Ala Ile 210 215 220		
Tyr Val Gly Leu Ile Thr Pro Ala Val His Phe Thr Met Gly Gly Val 225 230 235 240		
Arg Ile Asn Ser Arg Ala Glu Val Leu Ser Ser Ala Gly Thr Thr Tyr 245 250 255		
Lys Gly Leu Tyr Ala Ala Gly Glu Val Ser Gly Gly Val His Gly Arg 260 265 270		
Asn Arg Leu Gly Gly Ser Ser Leu Leu Glu Cys Val Val Phe Gly Arg 275 280 285		
Gln Ala Ala Thr Ser Ile 290		

<210> 18

<211> 2754

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (264)..(1772)

<223>

<220>

<221> misc_feature

<223> Oligo 162

<400> 18

tcgtggacca tttgtcgcac tggcgtgatg tctgtaaaag gaacggttac cgggatcacg	60
ccataaccagc tgctcccacg gctcgaaaag tagtaciaaat aggtctatag cgacggtgtg	120
cggcgggcac tgctgatttc ggaattcgct gtagtcaacg gcgctggata caggggcacc	180
attcaggata cccagattta caggaaagct agtcgtatgg acaacttcgc ccgtgcagca	240
gtatatatttat ttgcatgtga aca atg tta aga ggg cgt agc aag aca gtg ctg	293
Met Leu Arg Gly Arg Ser Lys Thr Val Leu	
1 5 10	
aca ttg ata ctg tta ctt gta acg gtc aga ctc ctg ttt aat cgc cta	341
Thr Leu Ile Leu Leu Leu Val Thr Val Arg Leu Leu Phe Asn Arg Leu	
15 20 25	
cgt tcg aga ggc cca gaa agc aat cgg agg ata gac atg tct ggt aaa	389
Arg Ser Arg Gly Pro Glu Ser Asn Arg Arg Ile Asp Met Ser Gly Lys	
30 35 40	
gtt cag cct gtt gtc gtc gtc gga ctg ggt ctt gcc ggt ctg tcg act	437
Val Gln Pro Val Val Val Val Gly Leu Gly Leu Ala Gly Leu Ser Thr	
45 50 55	
ggg gcg cag cta gtc aag aat ggg gtg ccg gtg atc ttg atg gac aag	485
Gly Ala Gln Leu Val Lys Asn Gly Val Pro Val Ile Leu Met Asp Lys	
60 65 70	
gtt tct gcc atc ggc gga aat tcc gtt aag gcc tcc agc ggc ata aac	533
Val Ser Ala Ile Gly Gly Asn Ser Val Lys Ala Ser Ser Gly Ile Asn	
75 80 85 90	
ggc gat ggt act cag gtg cag gag ctt ctt ggt gtg tat gac tct gct	581
Gly Asp Gly Thr Gln Val Gln Glu Leu Leu Gly Val Tyr Asp Ser Ala	
95 100 105	
gac tcc ttc tac aga gac acc gtt gcg tcg gcg aag ggt gcc ggc gcc	629
Asp Ser Phe Tyr Arg Asp Thr Val Ala Ser Ala Lys Gly Ala Gly Ala	
110 115 120	
agc gag cta atg gac aag ctt gca agg gac tcc gca ggt gct gta tcg	677
Ser Glu Leu Met Asp Lys Leu Ala Arg Asp Ser Ala Gly Ala Val Ser	
125 130 135	

tgg ctg cag aat gaa ttt gga gtt aaa ctg gac atc ctt tcg caa ctg	725
Trp Leu Gln Asn Glu Phe Gly Val Lys Leu Asp Ile Leu Ser Gln Leu	
140 145 150	
gga ggg cat tcg aag gct agg aca cat cgg tcg agc agc aac ata ccc	773
Gly Gly His Ser Lys Ala Arg Thr His Arg Ser Ser Ser Asn Ile Pro	
155 160 165 170	
cca ggg ttt gag atc ata tcg gcc atg cgc aag aga ctg gag tca gtg	821
Pro Gly Phe Glu Ile Ile Ser Ala Met Arg Lys Arg Leu Glu Ser Val	
175 180 185	
cag aag gaa aac cca gct ctc gtg aac ttc cgc ttg gag agc aaa gtt	869
Gln Lys Glu Asn Pro Ala Leu Val Asn Phe Arg Leu Glu Ser Lys Val	
190 195 200	
gtt gat ctc gaa ctg gcg gat gat gca gtc caa gga cta cgc tac atg	917
Val Asp Leu Glu Leu Ala Asp Asp Ala Val Gln Gly Leu Arg Tyr Met	
205 210 215	
gac aag gcc ggg gca cag cac ata ttg cac acg aag aat gta gtt ttc	965
Asp Lys Ala Gly Ala Gln His Ile Leu His Thr Lys Asn Val Val Phe	
220 225 230	
tgt act ggc ggg ttt ggg tcg tcc aag gag ttg ttg ggg aag tat gcc	1013
Cys Thr Gly Gly Phe Gly Ser Ser Lys Glu Leu Leu Gly Lys Tyr Ala	
235 240 245 250	
ccc cat ctc cgg gac ctt cct aca aca aac gcg cag ggt act acg gga	1061
Pro His Leu Arg Asp Leu Pro Thr Thr Asn Ala Gln Gly Thr Thr Gly	
255 260 265	
gat gga cag gac ctg atg ctc cgt gtt ggt ggg cag ctg att gat atg	1109
Asp Gly Gln Asp Leu Met Leu Arg Val Gly Gly Gln Leu Ile Asp Met	
270 275 280	
gag cat atc cag gta cac ccc acg agt ttt gtg gat cct aaa gat aaa	1157
Glu His Ile Gln Val His Pro Thr Ser Phe Val Asp Pro Lys Asp Lys	
285 290 295	
gac agc ggt agc aaa ttc ctg gcg gct gag gca ctg cgc ggc aat gga	1205
Asp Ser Gly Ser Lys Phe Leu Ala Ala Glu Ala Leu Arg Gly Asn Gly	
300 305 310	
ggt gtg ctt tta aac ccc agc act ggc cgg cgc ttt acg aac gaa ttg	1253
Gly Val Leu Leu Asn Pro Ser Thr Gly Arg Arg Phe Thr Asn Glu Leu	
315 320 325 330	
gcg aca cgt gat gta ctt acg gct agc att caa gag cac tgc aaa gag	1301
Ala Thr Arg Asp Val Leu Thr Ala Ser Ile Gln Glu His Cys Lys Glu	
335 340 345	
cat gtg gct tac ctt gtt ctg agc caa ggg gca tat gaa act ttg caa	1349
His Val Ala Tyr Leu Val Leu Ser Gln Gly Ala Tyr Glu Thr Leu Gln	
350 355 360	
agc ttt gtc aac ttc tac atc tct aaa ggg ctg atg cga gcg tcc acc	1397
Ser Phe Val Asn Phe Tyr Ile Ser Lys Gly Leu Met Arg Ala Ser Thr	
365 370 375	
gtt ggc gag ctt tcc gac gag atc gga gtt ccg aaa caa cag ctg gcg	1445

Val	Gly	Glu	Leu	Ser	Asp	Glu	Ile	Gly	Val	Pro	Lys	Gln	Gln	Leu	Ala		
380						385					390						
tcg	gag	cta	gag	gcg	tac	tcc	gcc	gca	gag	gtg	gat	caa	ttc	cag	cgt	1493	
Ser	Glu	Leu	Glu	Ala	Tyr	Ser	Ala	Ala	Glu	Val	Asp	Gln	Phe	Gln	Arg		
395					400					405					410		
cgg	cac	att	atc	aat	gat	ttt	ggc	ccg	agt	gtt	gat	ggg	tcg	aca	gca	1541	
Arg	His	Ile	Ile	Asn	Asp	Phe	Gly	Pro	Ser	Val	Asp	Gly	Ser	Thr	Ala		
				415					420					425			
ata	tat	gtt	ggc	ttg	atc	aca	ccc	gca	gtc	cat	ttt	acc	atg	ggc	ggc	1589	
Ile	Tyr	Val	Gly	Leu	Ile	Thr	Pro	Ala	Val	His	Phe	Thr	Met	Gly	Gly		
			430					435					440				
gtt	cga	atc	aat	tcc	aga	gcg	gaa	gtg	ctg	agc	agc	gca	ggg	acg	acc	1637	
Val	Arg	Ile	Asn	Ser	Arg	Ala	Glu	Val	Leu	Ser	Ser	Ala	Gly	Thr	Thr		
			445				450					455					
tat	aaa	ggg	ctt	tac	gcc	gct	gga	gag	gta	tcc	ggc	ggc	gtg	cac	ggc	1685	
Tyr	Lys	Gly	Leu	Tyr	Ala	Ala	Gly	Glu	Val	Ser	Gly	Gly	Val	His	Gly		
			460				465				470						
aga	aac	agg	ctg	ggc	ggg	tct	agt	ctc	ctg	gag	tgc	gtc	gtt	ttc	ggg	1733	
Arg	Asn	Arg	Leu	Gly	Gly	Ser	Ser	Leu	Leu	Glu	Cys	Val	Val	Phe	Gly		
475					480					485					490		
agg	cag	gca	gca	aca	tcg	att	aca	gaa	agg	ttg	cag	gcg	tgaactctag			1782	
Arg	Gln	Ala	Ala	Thr	Ser	Ile	Thr	Glu	Arg	Leu	Gln	Ala					
			495					500									
caaccctggc tcagtaacag acagtttttg cctgaaatgt gtcattttaaa taacagcaga																	1842
tagtaggtag catacaaaat taattagaga cgtcaattaa tactgcaatt aggaaactag																	1902
ggatatcaac tataattatt aaaaacgtcg agtaaaaaga attaatggga cggtccttga																	1962
aatagagccg aagaagctga agaagaacag taaatagtga atacgatgag aaaggccttc																	2022
aggtgcccgt tttgagtggg tagacataaa gcatttttgt ttcttagatc aatcgaaagg																	2082
aatgtgcgtc cttggatatt ttgatgctct ttttaggtac aggtggactg ccaaaaaatg																	2142
gcttgttcgc tagcctaatz ggattatttg ctgaccaat ctgccgtagt cagttagagg																	2202
gattcgteat ccacagtgtt cttagttttc ctaatcttga aaaacttcac accgctctgg																	2262
tgccgctgag gctctgcgac agggggcctgc accggggcgca cgtcattgac gggcacagcg																	2322
ccgcccetcaa gctcgggggt cgtgtcgtac acaatcgagt ggtccatctg gccatcgctc																	2382
ccgtgcttgg tgccctggta cttgagcgcc tgctcettgt agtgctgcat ctcaacgtac																	2442
tcottgtgtg cggccccagc ggggaggatg aaggagatcc cgctgaagaa gaagaccagc																	2502
agcgagcacg ccacctggc ccacgcaate cecatcaget ccggcccgat cttegcectg																	2562
cgcccgctcgt cgcggaacac gttgcgcgcc atcacaacaa ccgccgtctg gcacgccact																	2622
gtcgctacgt tgaacacgca cccgatcagc acctagggtga aaccaccttg gtgaacgagt																	2682
tcgaqcacca cgtgaacaca cagaacaaaa aacgcaattc gcaatgaaac gagagcgaga																	2742

tccagaagag ca

2754

<210> 19

<211> 503

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 162

<400> 19

Met Leu Arg Gly Arg Ser Lys Thr Val Leu Thr Leu Ile Leu Leu Leu
 1 5 10 15

Val Thr Val Arg Leu Leu Phe Asn Arg Leu Arg Ser Arg Gly Pro Glu
 20 25 30

Ser Asn Arg Arg Ile Asp Met Ser Gly Lys Val Gln Pro Val Val Val
 35 40 45

Val Gly Leu Gly Leu Ala Gly Leu Ser Thr Gly Ala Gln Leu Val Lys
 50 55 60

Asn Gly Val Pro Val Ile Leu Met Asp Lys Val Ser Ala Ile Gly Gly
 65 70 75 80

Asn Ser Val Lys Ala Ser Ser Gly Ile Asn Gly Asp Gly Thr Gln Val
 85 90 95

Gln Glu Leu Leu Gly Val Tyr Asp Ser Ala Asp Ser Phe Tyr Arg Asp
 100 105 110

Thr Val Ala Ser Ala Lys Gly Ala Gly Ala Ser Glu Leu Met Asp Lys
 115 120 125

Leu Ala Arg Asp Ser Ala Gly Ala Val Ser Trp Leu Gln Asn Glu Phe
 130 135 140

Gly Val Lys Leu Asp Ile Leu Ser Gln Leu Gly Gly His Ser Lys Ala
 145 150 155 160

Arg Thr His Arg Ser Ser Ser Asn Ile Pro Pro Gly Phe Glu Ile Ile
 165 170 175

Ser Ala Met Arg Lys Arg Leu Glu Ser Val Gln Lys Glu Asn Pro Ala
 180 185 190

Leu Val Asn Phe Arg Leu Glu Ser Lys Val Val Asp Leu Glu Leu Ala
 195 200 205

Asp Asp Ala Val Gln Gly Leu Arg Tyr Met Asp Lys Ala Gly Ala Gln
 210 215 220

His Ile Leu His Thr Lys Asn Val Val Phe Cys Thr Gly Gly Phe Gly
 225 230 235 240

Ser Ser Lys Glu Leu Leu Gly Lys Tyr Ala Pro His Leu Arg Asp Leu
 245 250 255

Pro Thr Thr Asn Ala Gln Gly Thr Thr Gly Asp Gly Gln Asp Leu Met
 260 265 270

Leu Arg Val Gly Gly Gln Leu Ile Asp Met Glu His Ile Gln Val His
 275 280 285

Pro Thr Ser Phe Val Asp Pro Lys Asp Lys Asp Ser Gly Ser Lys Phe
 290 295 300

Leu Ala Ala Glu Ala Leu Arg Gly Asn Gly Gly Val Leu Leu Asn Pro
 305 310 315 320

Ser Thr Gly Arg Arg Phe Thr Asn Glu Leu Ala Thr Arg Asp Val Leu
 325 330 335

Thr Ala Ser Ile Gln Glu His Cys Lys Glu His Val Ala Tyr Leu Val
 340 345 350

Leu Ser Gln Gly Ala Tyr Glu Thr Leu Gln Ser Phe Val Asn Phe Tyr
 355 360 365

Ile Ser Lys Gly Leu Met Arg Ala Ser Thr Val Gly Glu Leu Ser Asp
 370 375 380

Glu Ile Gly Val Pro Lys Gln Gln Leu Ala Ser Glu Leu Glu Ala Tyr
 385 390 395 400

Ser Ala Ala Glu Val Asp Gln Phe Gln Arg Arg His Ile Ile Asn Asp
 405 410 415

Phe Gly Pro Ser Val Asp Gly Ser Thr Ala Ile Tyr Val Gly Leu Ile
 420 425 430

Thr Pro Ala Val His Phe Thr Met Gly Gly Val Arg Ile Asn Ser Arg
 435 440 445

Ala Glu Val Leu Ser Ser Ala Gly Thr Thr Tyr Lys Gly Leu Tyr Ala
 450 455 460

Ala Gly Glu Val Ser Gly Gly Val His Gly Arg Asn Arg Leu Gly Gly
 465 470 475 480

Ser Ser Leu Leu Glu Cys Val Val Phe Gly Arg Gln Ala Ala Thr Ser
 485 490 495

Ile Thr Glu Arg Leu Gln Ala
 500

<210> 20

<211> 1060

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 178

<400> 20

gatcaatttc cggtgctggt ccgcagacaa ggctgtctta ccagtgccgc tcaagccgaa	60
gaagagcgtc acatcgttgt tggggccctg gttggcggag gagtgcaatg tcaagacgtt	120
gtgggtgata ggcatcaagt agaacatcac cgtgaagata cccttcttca tctcaccagc	180
gtactccgtg ccaaggatga ccatctccat cgccctgaag ttgatttcga cagtgggtctt	240
cgaggacata ccctcagtgt gcgtgttcgc tgggaactgg cccgcgttcc acacgggtgaa	300
gtctggctcg ccgaactccg caagctcctc ttcggtcggc ctgatcaaca tggtggatcat	360
gaacaacgcg tggatatgcg ggcagcacac aatccggatc ttgatgcggt agcggggggtc	420
ccagcccgcg tacgcatcca cgatgtatag gtgctcgcgg gtacgcaaga agtctgcggc	480
acgctcgcgg ttgatcaacc acgtcttctc cgacgccttg cggttcactg gacccaccca	540

```

aatattgtca gacgacgtgg gctcgtccac gatccgcttg tccttaggcg accgaccogt    600
cttttcccca gagtacgcaa tcaacgcgcc agtgctggag attgctgtct tgcgctccgc    660
tagcgcatct tcgtacaaca ccgcggccgg cgcattgcgc cgaatcgta aaacgcccga    720

gctgagccca agctcagcgc ggatcttctc ctctggtggt tggtatacct ttgatggcga    780
catagtgaac ttattggcct gctagctctc ctctgtgtag ctgaagcttg tccacttgag    840
acatgccatc tactacgcta gatgcctcta tttatacacg ccgaagcaca aggactgaaa    900
attacgtcta cttggctggt tgccgcagcg cccagtctct ggccgttcca ctgtgttcta    960
cgcggcctgc aacagcgctg tcgctagcac aaactacgtc gcggacctat ggatgaagca   1020
agtacagcgg tttcgctgtg ctatccgttt ttcgccgata                           1060

```

<210> 21

<211> 261

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 178

<400> 21

```

Met Ser Pro Ser Lys Val Tyr Gln Thr Pro Glu Glu Lys Ile Arg Ala
1           5           10           15

```

```

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

```

```

Ala Ala Val Leu Tyr Glu Asp Ala Leu Ala Glu Arg Lys Thr Ala Ile
35           40           45

```

```

Ser Ser Thr Gly Ala Leu Ile Ala Tyr Ser Gly Glu Lys Thr Gly Arg
50           55           60

```

```

Ser Pro Lys Asp Lys Arg Ile Val Asp Glu Pro Thr Ser Ser Asp Asn
65           70           75           80

```

```

Ile Trp Trp Gly Pro Val Asn Arg Lys Ala Ser Glu Lys Thr Trp Leu

```

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

<210> 22

<211> 2240

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (394) .. (2016)

<223>

<220>

<221> misc_feature

<223> Oligo 178

<400> 22

```

ccagcacaaa aaagagccgc gctgctccgg gcagccgccg gcagacggag ccgagcactt      60
accgaaaggc ctttccaaca gctgctaaat gtggggggatt gattcgcattg cttccggatc    120
ggcgaaaaaac ggatagcaca gcgaaaccgc tgtacttgct tcattccatag gtccgcgacg    180
tagttttgtgc tagcgacagc gctggttcag gccgcgtaga acacagtgga acggccagag    240
actgggcgct gcggcaaaca gccaaagtaga cgtaattttc agtccttggtg cttcggcggtg    300
tataaataga ggcatctagc gtagtagatg gcatgtotca agtggacaag cttcagctac      360
acagaggaga gctagcaggc caataagttc act atg tcg cca tca aag gta tac      414
                               Met Ser Pro Ser Lys Val Tyr
                               1             5

caa aca cca gag gag aag atc cgc gct gag ctt ggg ctc agc tcg ggc      462
Gln Thr Pro Glu Glu Lys Ile Arg Ala Glu Leu Gly Leu Ser Ser Gly
              10              15              20

gtt ttg acg att cgg cgc aat gcg ccg gcc gcg gtg ttg tac gaa gat      510
Val Leu Thr Ile Arg Arg Asn Ala Pro Ala Ala Val Leu Tyr Glu Asp
              25              30              35

gcg cta gcg gag cgc aag aca gca atc tcc agc act ggc gcg ttg att      558
Ala Leu Ala Glu Arg Lys Thr Ala Ile Ser Ser Thr Gly Ala Leu Ile
              40              45              50              55

gcg tac tct ggg gaa aag acg ggt cgg tcg cct aag gac aag cgg atc      606
Ala Tyr Ser Gly Glu Lys Thr Gly Arg Ser Pro Lys Asp Lys Arg Ile
              60              65              70

gtg gac gag ccc acg tcg tct gac aat att tgg tgg ggt cca gtg aac      654
Val Asp Glu Pro Thr Ser Ser Asp Asn Ile Trp Trp Gly Pro Val Asn
              75              80              85

cgc aag gcg tcg gag aag acg tgg ttg atc aac cgc gag cgt gcc gca      702
Arg Lys Ala Ser Glu Lys Thr Trp Leu Ile Asn Arg Glu Arg Ala Ala
              90              95              100

gac ttc ttg cgt acc cgc gag cac cta tac atc gtg gat gcg tac gcg      750
Asp Phe Leu Arg Thr Arg Glu His Leu Tyr Ile Val Asp Ala Tyr Ala
              105              110              115

ggc tgg gac ccc cgc tac cgc atc aag atc cgg att gtg tgc tgc cgc      798
Gly Trp Asp Pro Arg Tyr Arg Ile Lys Ile Arg Ile Val Cys Cys Arg
              120              125              130              135

```

gca tac cac gcg ttg ttc atg acc aac atg ttg atc agg ccg acc gaa	846
Ala Tyr His Ala Leu Phe Met Thr Asn Met Leu Ile Arg Pro Thr Glu	
140 145 150	
gag gag ctt gcg gag ttc ggc gag cca gac ttc acc gtg tgg aac gcg	894
Glu Glu Leu Ala Glu Phe Gly Glu Pro Asp Phe Thr Val Trp Asn Ala	
155 160 165	
ggc cag ttc cca gcg aac acg cac act gag ggt atg tcc tcg aag acc	942
Gly Gln Phe Pro Ala Asn Thr His Thr Glu Gly Met Ser Ser Lys Thr	
170 175 180	
act gtc gaa atc aac ttc agg gcg atg gag atg gtc atc ctt ggc acg	990
Thr Val Glu Ile Asn Phe Arg Ala Met Glu Met Val Ile Leu Gly Thr	
185 190 195	
gag tac gct ggt gag atg aag aag ggt atc ttc acg gtg atg ttc tac	1038
Glu Tyr Ala Gly Glu Met Lys Lys Gly Ile Phe Thr Val Met Phe Tyr	
200 205 210 215	
ttg atg cct atc aac cac aac gtc ttg aca ttg cac tcc tcc gcc aac	1086
Leu Met Pro Ile Asn His Asn Val Leu Thr Leu His Ser Ser Ala Asn	
220 225 230	
cag ggc ccc aac agc gat gtg acg ctc ttc ttc ggc ttg agc ggc act	1134
Gln Gly Pro Asn Ser Asp Val Thr Leu Phe Phe Gly Leu Ser Gly Thr	
235 240 245	
ggt aag acg acc ttg tct gcg gac cag cac cgg aaa ttg atc ggc gac	1182
Gly Lys Thr Thr Leu Ser Ala Asp Gln His Arg Lys Leu Ile Gly Asp	
250 255 260	
gac gag cac tgc tgg tct gac tac ggt gtg ttc aac ata gaa ggt ggg	1230
Asp Glu His Cys Trp Ser Asp Tyr Gly Val Phe Asn Ile Glu Gly Gly	
265 270 275	
tgc tac gcc aag tgt atc ggc cta agc ggg gag aag gag ccg gaa atc	1278
Cys Tyr Ala Lys Cys Ile Gly Leu Ser Gly Glu Lys Glu Pro Glu Ile	
280 285 290 295	
ttc aat gca atc aag tac ggc tct gtt ttg gag aac gtg gtc tac gac	1326
Phe Asn Ala Ile Lys Tyr Gly Ser Val Leu Glu Asn Val Val Tyr Asp	
300 305 310	
cca gta acg cgt gtc gtg gac tat gag gac tct tcg atc acc gag aac	1374
Pro Val Thr Arg Val Val Asp Tyr Glu Asp Ser Ser Ile Thr Glu Asn	
315 320 325	
aca cgt tgc gcg tac cca att gag tac atc cca agc gcg cag att cca	1422
Thr Arg Cys Ala Tyr Pro Ile Glu Tyr Ile Pro Ser Ala Gln Ile Pro	
330 335 340	
tgc ttg tcc gag aac cat cct tct aac atc gtt cta ttg acc tgc gat	1470
Cys Leu Ser Glu Asn His Pro Ser Asn Ile Val Leu Leu Thr Cys Asp	
345 350 355	
gcc tca ggt gtg ttg cca cct gtg tcc cgt ttg acg ccg gag cag gtg	1518
Ala Ser Gly Val Leu Pro Pro Val Ser Arg Leu Thr Pro Glu Gln Val	
360 365 370 375	
atg tac cac ttc atc tcg ggc tac acc tcc aag atg gcc ggt act gag	1566

Met	Tyr	His	Phe	Ile	Ser	Gly	Tyr	Thr	Ser	Lys	Met	Ala	Gly	Thr	Glu		
				380					385					390			
cag	gga	gtc	act	gag	cct	gag	gct	acc	ttc	tct	tcc	tgc	ttc	ggc	cag	1614	
Gln	Gly	Val	Thr	Glu	Pro	Glu	Ala	Thr	Phe	Ser	Ser	Cys	Phe	Gly	Gln		
				395					400					405			
cca	ttc	cta	gcc	cta	cat	cca	atg	aaa	tac	gct	acg	atg	ttg	gct	gaa	1662	
Pro	Phe	Leu	Ala	Leu	His	Pro	Met	Lys	Tyr	Ala	Thr	Met	Leu	Ala	Glu		
				410					415					420			
aag	atg	tcg	cag	cat	aac	gct	agc	gcc	tgg	ttg	att	aac	acg	ggc	tgg	1710	
Lys	Met	Ser	Gln	His	Asn	Ala	Ser	Ala	Trp	Leu	Ile	Asn	Thr	Gly	Trp		
				425					430					435			
act	ggc	tcc	tcg	tat	acg	gcc	ggc	ggt	aag	cgc	tgt	cca	ttg	aag	tac	1758	
Thr	Gly	Ser	Ser	Tyr	Thr	Ala	Gly	Gly	Lys	Arg	Cys	Pro	Leu	Lys	Tyr		
				440					445					450			455
acc	cgt	gct	atc	ttg	gac	tcg	atc	cat	gac	gga	acc	ttg	gcc	cag	gct	1806	
Thr	Arg	Ala	Ile	Leu	Asp	Ser	Ile	His	Asp	Gly	Thr	Leu	Ala	Gln	Ala		
				460					465					470			
gaa	tac	gag	acc	cta	ccc	gtc	ttt	ggt	ttg	tcc	att	cca	aaa	gcg	gtc	1854	
Glu	Tyr	Glu	Thr	Leu	Pro	Val	Phe	Gly	Leu	Ser	Ile	Pro	Lys	Ala	Val		
				475					480					485			
gag	ggc	gtg	cca	gct	gag	ttg	cta	aac	cct	gcc	aag	aac	tgg	gtc	gag	1902	
Glu	Gly	Val	Pro	Ala	Glu	Leu	Leu	Asn	Pro	Ala	Lys	Asn	Trp	Val	Glu		
				490					495					500			
ggc	gaa	ggc	aag	tac	gct	agc	gcc	gtg	tcg	gct	ttg	gcc	tcg	aag	ttc	1950	
Gly	Glu	Gly	Lys	Tyr	Ala	Ser	Ala	Val	Ser	Ala	Leu	Ala	Ser	Lys	Phe		
				505					510					515			
aca	gag	aac	ttt	aag	atc	tat	cag	gac	cag	gcc	aca	gag	gag	gtt	gtc	1998	
Thr	Glu	Asn	Phe	Lys	Ile	Tyr	Gln	Asp	Gln	Ala	Thr	Glu	Glu	Val	Val		
				520					525					530			535
cgc	gct	ggc	cca	caa	ata	tgagatgcga	ccatatagtg	gtttcagggga								2046	
Arg	Ala	Gly	Pro	Gln	Ile												
				540													
tctgagtcctt	ccggacccta	tttataagggt	aaaatactat	atgtaataaaa	taacatatctc											2106	
ttgagaaaact	gcctgcctca	tcggcgcggt	gtcactaggc	gcccaggggg	gccgtaacca											2166	
cagtttgcag	tcatgcgggg	tggtggcatc	gtagtcgtag	tggtgtgtgc	tgttccagca											2226	
gcaggcttct	cact														2240		
<210>	23																
<211>	541																
<212>	PRT																
<213>	Ashbya gossypii																

<220>

<221> misc_feature

<223> Oligo 178

<400> 23

Met Ser Pro Ser Lys Val Tyr Gln Thr Pro Glu Glu Lys Ile Arg Ala
1 5 10 15

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

Ala Ala Val Leu Tyr Glu Asp Ala Leu Ala Glu Arg Lys Thr Ala Ile
35 40 45

Ser Ser Thr Gly Ala Leu Ile Ala Tyr Ser Gly Glu Lys Thr Gly Arg
50 55 60

Ser Pro Lys Asp Lys Arg Ile Val Asp Glu Pro Thr Ser Ser Asp Asn
65 70 75 80

Ile Trp Trp Gly Pro Val Asn Arg Lys Ala Ser Glu Lys Thr Trp Leu
85 90 95

Ile Asn Arg Glu Arg Ala Ala Asp Phe Leu Arg Thr Arg Glu His Leu
100 105 110

Tyr Ile Val Asp Ala Tyr Ala Gly Trp Asp Pro Arg Tyr Arg Ile Lys
115 120 125

Ile Arg Ile Val Cys Cys Arg Ala Tyr His Ala Leu Phe Met Thr Asn
130 135 140

Met Leu Ile Arg Pro Thr Glu Glu Glu Leu Ala Glu Phe Gly Glu Pro
145 150 155 160

Asp Phe Thr Val Trp Asn Ala Gly Gln Phe Pro Ala Asn Thr His Thr
165 170 175

Glu Gly Met Ser Ser Lys Thr Thr Val Glu Ile Asn Phe Arg Ala Met
180 185 190

Glu Met Val Ile Leu Gly Thr Glu Tyr Ala Gly Glu Met Lys Lys Gly
195 200 205

Ile Phe Thr Val Met Phe Tyr Leu Met Pro Ile Asn His Asn Val Leu

210		215		220
Thr Leu His Ser Ser Ala Asn Gln Gly Pro Asn Ser Asp Val Thr Leu				
225		230		235 240
Phe Phe Gly Leu Ser Gly Thr Gly Lys Thr Thr Leu Ser Ala Asp Gln				
	245		250	255
His Arg Lys Leu Ile Gly Asp Asp Glu His Cys Trp Ser Asp Tyr Gly				
	260		265	270
Val Phe Asn Ile Glu Gly Gly Cys Tyr Ala Lys Cys Ile Gly Leu Ser				
	275		280	285
Gly Glu Lys Glu Pro Glu Ile Phe Asn Ala Ile Lys Tyr Gly Ser Val				
	290		295	300
Leu Glu Asn Val Val Tyr Asp Pro Val Thr Arg Val Val Asp Tyr Glu				
305		310		315 320
Asp Ser Ser Ile Thr Glu Asn Thr Arg Cys Ala Tyr Pro Ile Glu Tyr				
	325		330	335
Ile Pro Ser Ala Gln Ile Pro Cys Leu Ser Glu Asn His Pro Ser Asn				
	340		345	350
Ile Val Leu Leu Thr Cys Asp Ala Ser Gly Val Leu Pro Pro Val Ser				
	355		360	365
Arg Leu Thr Pro Glu Gln Val Met Tyr His Phe Ile Ser Gly Tyr Thr				
	370		375	380
Ser Lys Met Ala Gly Thr Glu Gln Gly Val Thr Glu Pro Glu Ala Thr				
385		390		395 400
Phe Ser Ser Cys Phe Gly Gln Pro Phe Leu Ala Leu His Pro Met Lys				
	405		410	415
Tyr Ala Thr Met Leu Ala Glu Lys Met Ser Gln His Asn Ala Ser Ala				
	420		425	430
Trp Leu Ile Asn Thr Gly Trp Thr Gly Ser Ser Tyr Thr Ala Gly Gly				
	435		440	445
Lys Arg Cys Pro Leu Lys Tyr Thr Arg Ala Ile Leu Asp Ser Ile His				
	450		455	460

Asp Gly Thr Leu Ala Gln Ala Glu Tyr Glu Thr Leu Pro Val Phe Gly
 465 470 475 480

Leu Ser Ile Pro Lys Ala Val Glu Gly Val Pro Ala Glu Leu Leu Asn
 485 490 495

Pro Ala Lys Asn Trp Val Glu Gly Glu Gly Lys Tyr Ala Ser Ala Val
 500 505 510

Ser Ala Leu Ala Ser Lys Phe Thr Glu Asn Phe Lys Ile Tyr Gln Asp
 515 520 525

Gln Ala Thr Glu Glu Val Val Arg Ala Gly Pro Gln Ile
 530 535 540

<210> 24

<211> 1058

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 64

<400> 24

```

tcaccaaaag ggtgctcaag tttgtcctcc ggaaccggaa gacagccttc gttagtttca      60
tctccggctg ggtggtttca ataacgtagt ctttggacaa tcaaaccctg taatctaccc      120
tggagcttcg gtttatgaag aggagagaga agtctagcgg caaaagaagc tcaatttaaa      180
atagagattdt gtgtttgcca ctacgatgtt cagccgcgca ttccagatac attcgtctgg      240
cacgatgaaa taaccgccga agtctcctca cgtgattatt acgtaacata ctttccgggt      300
ctcgaggatg cagtagtaca ccccgaaata atgctacatt agtcacgtgg agatactgca      360
gaagttgaag aagacaacgg ctttctccac tactacatct acgaacagca aggatcatgc      420
aagctgttga tcggacgcag tttgcgccac ttaagaatga cttgatgtta agggctgcga      480
ggggtgaaga tggtgaacga cctccgtgct ggatcatgcg gcaagcaggt cggacttac      540
ccgaatacca tgacgtgaag ggtggccgtg acttctttga aacctgccgt gatgctgaaa      600
tagcttcgga aatcacgctt cagcccgtaa ggagatacgc aggccttttg gacgctgcaa      660

```

```

ttatcttctc ggatattctg gtgattccgc aggcgatggg gatgcgcgta gagatgctgg      720
agggcaaggg cccgcacttc cccgagccac tccggaccac cgagcaactg atgaaggtgc      780
tagagtacga agtggatgtg ctgcgtgagc tcgactgggc tttccggggc atcacactaa      840
cgcgggcaaaa gctggccggc gatgtgcctc tottcgggctt ctgtggaggg ccctggacat      900
tgctagtcta catgaccgag ggaggtggct cgcgtctggt ccgctttgcc aaagaatgga      960
ttcacaaacg tccagaactt gcgcataaac tactgcagaa aatcactgac gttgcaattg     1020
agttttttggc gcagcagggt gtagctgggt gccagatc                               1058

```

<210> 25

<211> 206

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 64

<400> 25

```

Phe Ala Pro Leu Lys Asn Asp Leu Met Leu Arg Ala Ala Arg Gly Glu
1           5           10           15

```

```

Asp Val Glu Arg Pro Pro Cys Trp Ile Met Arg Gln Ala Gly Arg Tyr
          20           25           30

```

```

Leu Pro Glu Tyr His Asp Val Lys Gly Gly Arg Asp Phe Phe Glu Thr
35           40           45

```

```

Cys Arg Asp Ala Glu Ile Ala Ser Glu Ile Thr Leu Gln Pro Val Arg
50           55           60

```

```

Arg Tyr Ala Gly Leu Leu Asp Ala Ala Ile Ile Phe Ser Asp Ile Leu
65           70           75           80

```

```

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

```

```

Gly Pro His Phe Pro Glu Pro Leu Arg Thr Thr Glu Gln Leu Met Lys
          100          105          110

```

Val Leu Glu Tyr Glu Val Asp Val Leu Arg Glu Leu Asp Trp Ala Phe
 115 120 125

Arg Ala Ile Thr Leu Thr Arg Gln Lys Leu Ala Gly Asp Val Pro Leu
 130 135 140

Phe Gly Phe Cys Gly Gly Pro Trp Thr Leu Leu Val Tyr Met Thr Glu
 145 150 155 160

Gly Gly Gly Ser Arg Leu Phe Arg Phe Ala Lys Glu Trp Ile His Lys
 165 170 175

Arg Pro Glu Leu Ala His Lys Leu Leu Gln Lys Ile Thr Asp Val Ala
 180 185 190

Ile Glu Phe Leu Ala Gln Gln Val Val Ala Gly Cys Gln Ile
 195 200 205

<210> 26

<211> 1905

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (612) .. (1712)

<223>

<220>

<221> misc_feature

<223> Oligi 64

<400> 26
 cctttaccct atccaacaaa taatcgtgta ccttgtcatt tgcattgcgac gcgtacggat 60
 gcggtatttg tgtgatgatc tccaaatcca accaggcatc tgcaagcatc tgcgcataat 120
 tatagtctga tggcaatgta aagtgggtacc taaagtgggtc ccacacataa agagttgtga 180
 tcaccaaata tgtgatcacc aaaagggtgc tcaagtttgt cctccggaac cggaagacag 240

ccttcgtttag tttcatctcc ggctgggtgg tttcaataac gtagtctttg gacaatcaaa	300
cccgttaatc taccctggag cttcggttta tgaagaggag agagaagtct agcggcaaaa	360
gaagctcaat ttaaaataga gatttgtgtt tgccactacg atgttcagcc gcgcattcca	420
gatacattcg tctggcacga tgaataaacc gccgaagtct cctcacgtga ttattacgta	480
acatactttc cgggtctcga ggatgcagta gtacaccccg aaataatgct acattagtca	540
cgtggagata ctgcagaagt tgaagaagac aacggctttc tccactacta catctacgaa	600
cagcaaggat c atg caa gct gtt gat cgg acg cag ttt gcg cca ctt aag	650
Met Gln Ala Val Asp Arg Thr Gln Phe Ala Pro Leu Lys	
1 5 10	
aat gac ttg atg tta agg gct gcg agg ggt gaa gat gtt gaa cga cct	698
Asn Asp Leu Met Leu Arg Ala Ala Arg Gly Glu Asp Val Glu Arg Pro	
15 20 25	
ccg tgc tgg atc atg cgg caa gca ggt cgg tac tta ccc gaa tac cat	746
Pro Cys Trp Ile Met Arg Gln Ala Gly Arg Tyr Leu Pro Glu Tyr His	
30 35 40 45	
gac gtg aag ggt ggc cgt gac ttc ttt gaa acc tgc cgt gat gct gaa	794
Asp Val Lys Gly Gly Arg Asp Phe Phe Glu Thr Cys Arg Asp Ala Glu	
50 55 60	
ata gct tcg gaa atc acg ctt cag ccc gta agg aga tac gca ggc ctt	842
Ile Ala Ser Glu Ile Thr Leu Gln Pro Val Arg Arg Tyr Ala Gly Leu	
65 70 75	
ttg gac gct gca att atc ttc tcg gat att ctg gtg att ccg cag gcg	890
Leu Asp Ala Ala Ile Ile Phe Ser Asp Ile Leu Val Ile Pro Gln Ala	
80 85 90	
atg ggg atg cgc gta gag atg ctg gag ggc aag ggc ccg cac ttc ccc	938
Met Gly Met Arg Val Glu Met Leu Glu Gly Lys Gly Pro His Phe Pro	
95 100 105	
gag cca ctc cgg acc acc gag caa ctg atg aag gtg cta gag tac gaa	986
Glu Pro Leu Arg Thr Thr Glu Gln Leu Met Lys Val Leu Glu Tyr Glu	
110 115 120 125	
gtg gat gtg ctg cgt gag ctc gac tgg gct ttc cgg gcc atc aca cta	1034
Val Asp Val Leu Arg Glu Leu Asp Trp Ala Phe Arg Ala Ile Thr Leu	
130 135 140	
acg cgg caa aag ctg gcc ggc gat gtg cct ctc ttc ggc ttc tgt gga	1082
Thr Arg Gln Lys Leu Ala Gly Asp Val Pro Leu Phe Gly Phe Cys Gly	
145 150 155	
ggg ccc tgg aca ttg cta gtc tac atg acc gag gga ggt ggc tcg cgt	1130
Gly Pro Trp Thr Leu Leu Val Tyr Met Thr Glu Gly Gly Gly Ser Arg	
160 165 170	
ctg ttc cgc ttt gcc aaa gaa tgg att cac aaa cgt cca gaa ctt gcg	1178
Leu Phe Arg Phe Ala Lys Glu Trp Ile His Lys Arg Pro Glu Leu Ala	
175 180 185	
cat aaa cta ctg cag aaa atc act gac gtt gca att gag ttt ttg gcg	1226
His Lys Leu Leu Gln Lys Ile Thr Asp Val Ala Ile Glu Phe Leu Ala	

190	195	200	205	
cag cag gtt gaa gct ggt tgc cag atc ctg cag ttg ttc gaa agt tgg				1274
Gln Gln Val Glu Ala Gly Cys Gln Ile Leu Gln Leu Phe Glu Ser Trp	210	215	220	
ggt ggc gag ctg tcc ctt gcc gac ttg gac gaa ttt tgc ctg ccg tac				1322
Gly Gly Glu Leu Ser Leu Ala Asp Leu Asp Glu Phe Ser Leu Pro Tyr	225	230	235	
ttg aag caa att gcg gct gcc gtt ccg gca cgt tta cgc gaa tta ggc				1370
Leu Lys Gln Ile Ala Ala Ala Val Pro Ala Arg Leu Arg Glu Leu Gly	240	245	250	
att act gaa gat gtc cca atg acg gtt ttt gcc aag ggc tgc tgg tat				1418
Ile Thr Glu Asp Val Pro Met Thr Val Phe Ala Lys Gly Ser Trp Tyr	255	260	265	
gct cta gac aag ctc tgc gac gca ggc tat aac acc gtt tcc ctc gac				1466
Ala Leu Asp Lys Leu Cys Asp Ala Gly Tyr Asn Thr Val Ser Leu Asp	270	275	280	285
tgg ctg tgg gat cct gcc gag gcc gtt cgt atc aac aac ggc aga gtc				1514
Trp Leu Trp Asp Pro Ala Glu Ala Val Arg Ile Asn Asn Gly Arg Val	290	295	300	
acc cta cag ggc aat ctg gac ccg acc gtc atg tat ggt tcc gat gag				1562
Thr Leu Gln Gly Asn Leu Asp Pro Thr Val Met Tyr Gly Ser Asp Glu	305	310	315	
act atc acc aag aag act gag gca atg atc aaa gga ttc ggt ggt ggc				1610
Thr Ile Thr Lys Lys Thr Glu Ala Met Ile Lys Gly Phe Gly Gly Gly	320	325	330	
aaa aaa cag tac atc gtc aat ttc ggt cac gga acg cac cct ttc atg				1658
Lys Lys Gln Tyr Ile Val Asn Phe Gly His Gly Thr His Pro Phe Met	335	340	345	
gac cca gag aag ata cgc cat ttt ctt gag gaa tgc cac aga atc ggg				1706
Asp Pro Glu Lys Ile Arg His Phe Leu Glu Glu Cys His Arg Ile Gly	350	355	360	365
aag atg taaccttctg gaagtagaag agcacattat caggtcaata gcatcattat				1762
Lys Met				
gtattatattt agacgctagt tgaactccaa tatttttgat catgtgagcg tatgacgggc				1822
cactttacca cgctagcaa aaatgttcgc gggtgttcta ccgatcataa accacttcac				1882
cgttcggaga agcacttttg atc				1905
<210>	27			
<211>	367			
<212>	PRT			
<213>	Ashbya gossypii			

<220>

<221> misc_feature

<223> Oligi 64

<400> 27

Met Gln Ala Val Asp Arg Thr Gln Phe Ala Pro Leu Lys Asn Asp Leu
 1 5 10 15

Met Leu Arg Ala Ala Arg Gly Glu Asp Val Glu Arg Pro Pro Cys Trp
 20 25 30

Ile Met Arg Gln Ala Gly Arg Tyr Leu Pro Glu Tyr His Asp Val Lys
 35 40 45

Gly Gly Arg Asp Phe Phe Glu Thr Cys Arg Asp Ala Glu Ile Ala Ser
 50 55 60

Glu Ile Thr Leu Gln Pro Val Arg Arg Tyr Ala Gly Leu Leu Asp Ala
 65 70 75 80

Ala Ile Ile Phe Ser Asp Ile Leu Val Ile Pro Gln Ala Met Gly Met
 85 90 95

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

Arg Thr Thr Glu Gln Leu Met Lys Val Leu Glu Tyr Glu Val Asp Val
 115 120 125

Leu Arg Glu Leu Asp Trp Ala Phe Arg Ala Ile Thr Leu Thr Arg Gln
 130 135 140

Lys Leu Ala Gly Asp Val Pro Leu Phe Gly Phe Cys Gly Gly Pro Trp
 145 150 155 160

Thr Leu Leu Val Tyr Met Thr Glu Gly Gly Gly Ser Arg Leu Phe Arg
 165 170 175

Phe Ala Lys Glu Trp Ile His Lys Arg Pro Glu Leu Ala His Lys Leu
 180 185 190

Leu Gln Lys Ile Thr Asp Val Ala Ile Glu Phe Leu Ala Gln Gln Val
 195 200 205

Glu Ala Gly Cys Gln Ile Leu Gln Leu Phe Glu Ser Trp Gly Gly Glu
 210 215 220
 Leu Ser Leu Ala Asp Leu Asp Glu Phe Ser Leu Pro Tyr Leu Lys Gln
 225 230 235 240
 Ile Ala Ala Ala Val Pro Ala Arg Leu Arg Glu Leu Gly Ile Thr Glu
 245 250 255
 Asp Val Pro Met Thr Val Phe Ala Lys Gly Ser Trp Tyr Ala Leu Asp
 260 265 270
 Lys Leu Cys Asp Ala Gly Tyr Asn Thr Val Ser Leu Asp Trp Leu Trp
 275 280 285
 Asp Pro Ala Glu Ala Val Arg Ile Asn Asn Gly Arg Val Thr Leu Gln
 290 295 300
 Gly Asn Leu Asp Pro Thr Val Met Tyr Gly Ser Asp Glu Thr Ile Thr
 305 310 315 320
 Lys Lys Thr Glu Ala Met Ile Lys Gly Phe Gly Gly Gly Lys Lys Gln
 325 330 335
 Tyr Ile Val Asn Phe Gly His Gly Thr His Pro Phe Met Asp Pro Glu
 340 345 350
 Lys Ile Arg His Phe Leu Glu Glu Cys His Arg Ile Gly Lys Met
 355 360 365

<210> 28

<211> 972

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 125

<400> 28

atctgctgcg gccgatgct aacatccagg agcttggaat tctgcgtcgt gcgctggggcc 60
 tctagttttcc cgccgccgtt agtgagcaat atgaagggtta tgcagttggc aggtccagca 120

```

gcctgagcgc ctcggccgca cccggaatgg gcttttcgcc gtggaggagg acgccgtcga      180
tgtcaaacac aaacccgacg tctgtagaga ggcgcctgat acgctgtagg cacttgagtg      240
gcgagaagcg gagcattgct gtgctgggcg agctgccttc gcgaatgggg tctgtgttcc      300
gtcggcaact gcgttcgcct gcgaggcgga attttattgt gtgtcatatc tcgccggcgt      360
gagctcagcc tccgtcagcg acgtgcggca gagccacttg gcgcgtgaac tgggagcgct      420
atttactata caggaacaag tgttttctagt agtgctggga tgcatctccc ggtatccctc      480
ttcgacagag tagtgcaggg cgctgcctg ttccgacctt aaccctttta gtgaagtcac      540
tgctgaccg acaatcagca gccctggcgg gcgtgagccg atctgttcga cgacctcagg      600
cacgcttgca agcgtgggtt tggtcacgcg ctggtcggga cacgaggcac gctcgacaat      660
ggctgctggc acctgagggg cccattgctg ttccagcagg gcctggacaa ggtccccac      720
acgatgcagc gccattagga aaattgtagt gcgtgcggcg acaaactcgg gagcaggggg      780
cagcactccc cgctaccag gtgcctggta catatcaaga cctgctctgc gacgtcactg      840
ctgcgtggcc ggaatatgtg cgacagtagt ggcagcgaga gctgagctaa taccaggtac      900
aacgcgaggg tcaaaccocat gccgcttgaa aagcaggtac tcttcgccgc ctctaccgaa      960
gatgtacgga tc                                                    972

```

<210> 29

<211> 147

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 125

<400> 29

```

Asp Pro Tyr Ile Phe Gly Arg Gly Gly Glu Glu Tyr Leu Leu Phe Lys
1           5           10          15

```

```

Arg His Gly Phe Glu Pro Arg Val Val Pro Gly Ile Ser Ser Ala Leu
20          25          30

```

```

Ala Ala Thr Thr Val Ala His Ile Pro Ala Thr Gln Gln Arg Arg Arg
35          40          45

```

Ala Gly Leu Asp Met Tyr Gln Ala Pro Gly Arg Arg Gly Val Leu Pro
 50 55 60

Pro Ala Pro Glu Phe Val Ala Ala Arg Thr Thr Ile Phe Leu Met Ala
 65 70 75 80

Leu His Arg Val Gly Asp Leu Val Gln Ala Leu Leu Glu Gln Gln Trp
 85 90 95

Asp Pro Gln Val Pro Ala Ala Ile Val Glu Arg Ala Ser Cys Pro Asp
 100 105 110

Gln Arg Val Thr Arg Thr Thr Leu Ala Ser Val Pro Glu Val Val Glu
 115 120 125

Gln Ile Gly Ser Arg Pro Pro Gly Leu Leu Ile Val Gly His Ala Val
 130 135 140

Thr Ser Leu
 145

<210> 30

<211> 2455

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (275) .. (2023)

<223>

<220>

<221> misc_feature

<223> Oligo 125

<400> 30
 ggtcgcgaatg gaataataag gttgtatatata gttattctgc cccaattcga tgccctgagg 60
 ttttcggcga tgagcagtaa aaaagctcga tcaggccatc agcacgtgac tagtacaaag 120

ccacagctgt ggggttggtc aagcttatga tcagcgccgc tcaaggatga atcacaagcg	180
actgtggcca gacctataa agagatatat acgcccggga agatggaggg cgttttgtgg	240
ggtccatgct gtgattcttt aaaggggtgac atcg atg ggc cgt gaa gga gtg ccg	295
Met Gly Arg Glu Gly Val Pro	
1 5	
cta att gcg tct ttg aac gct agc aat gaa gtc atg cta ctg gtg ggg	343
Leu Ile Ala Ser Leu Asn Ala Ser Asn Glu Val Met Leu Leu Val Gly	
10 15 20	
aca ggc gga tca ggg cga gtg gtc agt cga ctg agg ggg ttg ctt aat	391
Thr Gly Gly Ser Gly Arg Val Val Ser Arg Leu Arg Gly Leu Leu Asn	
25 30 35	
aca ggt gca cag tgc att gtg gta cag ccc tcg aac agc aaa ctc gtg	439
Thr Gly Ala Gln Cys Ile Val Val Gln Pro Ser Asn Ser Lys Leu Val	
40 45 50 55	
gcg caa cta tat cgg gaa ttt gcg ggg gag ccg ctg ctc cgg atc tta	487
Ala Gln Leu Tyr Arg Glu Phe Ala Gly Glu Pro Leu Leu Arg Ile Leu	
60 65 70	
gat cgg gcg ttt gca ttg gaa gac ctg acg acc cta ggg agg tac gag	535
Asp Arg Ala Phe Ala Leu Glu Asp Leu Thr Thr Leu Gly Arg Tyr Glu	
75 80 85	
gcc agc aat gtg gtg gat cgg gta ttt ctg acg ccg ggt cac agt gct	583
Ala Ser Asn Val Val Asp Arg Val Phe Leu Thr Pro Gly His Ser Ala	
90 95 100	
gat gag ata tta cag gcc cat gcg gtg tac cgg cag tgc atc aaa ctg	631
Asp Glu Ile Leu Gln Ala His Ala Val Tyr Arg Gln Cys Ile Lys Leu	
105 110 115	
cgc att ccc atc aac acg agc caa cgg ccg gaa ctt tcg acc ttt tcg	679
Arg Ile Pro Ile Asn Thr Ser Gln Arg Pro Glu Leu Ser Thr Phe Ser	
120 125 130 135	
ctg ccg tca acc tac agg gac ccg aaa ggc tcc ggg ttg cag gtg gcc	727
Leu Pro Ser Thr Tyr Arg Asp Pro Lys Gly Ser Gly Leu Gln Val Ala	
140 145 150	
gtc act act gat ggc cgc ggc tgc gta ctg gcc aac aga atc aga cgt	775
Val Thr Thr Asp Gly Arg Gly Cys Val Leu Ala Asn Arg Ile Arg Arg	
155 160 165	
gaa att gtg agc aag ctg cct gcc aat atc tcc gag gta gtc agc aac	823
Glu Ile Val Ser Lys Leu Pro Ala Asn Ile Ser Glu Val Val Ser Asn	
170 175 180	
atg ggt cgg ctg cgg gac tgc ata acc aca cgt gac cat gcg gaa ttg	871
Met Gly Arg Leu Arg Asp Cys Ile Thr Thr Arg Asp His Ala Glu Leu	
185 190 195	
ctc cag acg gtt gac tta gag tcc gag ctc gag acc atg ggc atc ctg	919
Leu Gln Thr Val Asp Leu Glu Ser Glu Leu Glu Thr Met Gly Ile Leu	
200 205 210 215	
ccc gat gag gac caa tgg gcg tcg cac cgc ttt aat aag ctg gtc cag	967

Pro Asp Glu Asp Gln Trp Ala Ser His Arg Phe Asn Lys Leu Val Gln			
220	225	230	
gag ttt cgg atg aag gag tcc gac cag cga ctc aag cgg ctc cgg tgg	1015		
Glu Phe Arg Met Lys Glu Ser Asp Gln Arg Leu Lys Arg Leu Arg Trp			
235	240	245	
ctc tcc cag ata atg gag tac tac cct ttg agc gag cta gcg gat gtg	1063		
Leu Ser Gln Ile Met Glu Tyr Tyr Pro Leu Ser Glu Leu Ala Asp Val			
250	255	260	
tcc gtt gac aaa ctg ttg gat gag cag act gga gcc gcg tgt tct gga	1111		
Ser Val Asp Lys Leu Leu Asp Glu Gln Thr Gly Ala Ala Cys Ser Gly			
265	270	275	
acg aca gcg aca acc gga gcc gca tcc gtg act ggg gcc aca gag gac	1159		
Thr Thr Ala Thr Thr Gly Ala Ala Ser Val Thr Gly Ala Thr Glu Asp			
280	285	290	295
acg gct gct gag cct gct gag ggg gtg gca tcc gat agc tct ggc gct	1207		
Thr Ala Ala Glu Pro Ala Glu Gly Val Ala Ser Asp Ser Ser Gly Ala			
300	305	310	
cgc gac cta gca cgc agt gcc gca ggc tcc atc gca ctt gtt ggc agt	1255		
Arg Asp Leu Ala Arg Ser Ala Ala Gly Ser Ile Ala Leu Val Gly Ser			
315	320	325	
ggc cct ggc tcc ctc tcc atg ctc acg cta ggc gca ctc agc gag att	1303		
Gly Pro Gly Ser Leu Ser Met Leu Thr Leu Gly Ala Leu Ser Glu Ile			
330	335	340	
aag tcc gca gac ctt gtc ctc gca gac aag ctg gtt ccg caa gag gtg	1351		
Lys Ser Ala Asp Leu Val Leu Ala Asp Lys Leu Val Pro Gln Glu Val			
345	350	355	
atg aat aca att cca gcg ggc act gag acc ttc atc gcc aga aag ttc	1399		
Met Asn Thr Ile Pro Ala Gly Thr Glu Thr Phe Ile Ala Arg Lys Phe			
360	365	370	375
ccg ggc aat gcg gag cgc gca cag gaa gag cta gcc gaa aag gct cta	1447		
Pro Gly Asn Ala Glu Arg Ala Gln Glu Leu Ala Glu Lys Ala Leu			
380	385	390	
gcc gcg ctt cgg gag gga aag agg gtt gtg cgc ttg aaa cag ggc gat	1495		
Ala Ala Leu Arg Glu Gly Lys Arg Val Val Arg Leu Lys Gln Gly Asp			
395	400	405	
ccg tac atc ttc ggt aga ggc ggc gaa gag tac ctg ctt ttc aag cgg	1543		
Pro Tyr Ile Phe Gly Arg Gly Gly Glu Glu Tyr Leu Leu Phe Lys Arg			
410	415	420	
cat ggg ttt gag cct cgc gtt gta cct ggt att agc tca gct ctc gct	1591		
His Gly Phe Glu Pro Arg Val Val Pro Gly Ile Ser Ser Ala Leu Ala			
425	430	435	
gcc act act gtc gca cat att ccg gcc acg cag cgt gac gtc gca gag	1639		
Ala Thr Thr Val Ala His Ile Pro Ala Thr Gln Arg Asp Val Ala Glu			
440	445	450	455
cag gtc ttg ata tgt acc ggc act ggt agg cgg gga gtg ctg ccc cct	1687		
Gln Val Leu Ile Cys Thr Gly Thr Gly Arg Arg Gly Val Leu Pro Pro			
460	465	470	

```

gct ccc gag ttt gtc gcc gca cgc act aca att ttc cta atg gcg ctg      1735
Ala Pro Glu Phe Val Ala Ala Arg Thr Thr Ile Phe Leu Met Ala Leu
              475                      480                      485

cat cgt gtg ggg gac ctt gtc cag gcc ctg ctg gaa cag caa tgg gac      1783
His Arg Val Gly Asp Leu Val Gln Ala Leu Leu Glu Gln Gln Trp Asp
              490                      495                      500

cct cag gtg cca gca gcc att gtc gag cgt gcc tcg tgt ccg gac cag      1831
Pro Gln Val Pro Ala Ala Ile Val Glu Arg Ala Ser Cys Pro Asp Gln
              505                      510                      515

cgc gtg acc aga acc acg ctt gca agc gtg cct gag gtc gtc gaa cag      1879
Arg Val Thr Arg Thr Thr Leu Ala Ser Val Pro Glu Val Val Glu Gln
520                      525                      530                      535

atc ggc tca cgc ccg cca ggg ctg ctg att gtc ggt cac gca gtg act      1927
Ile Gly Ser Arg Pro Pro Gly Leu Leu Ile Val Gly His Ala Val Thr
              540                      545                      550

tca cta aaa ggg tta ggg tcc gaa cag gca ggc gcc ctg cac tac tct      1975
Ser Leu Lys Gly Leu Gly Ser Glu Gln Ala Gly Ala Leu His Tyr Ser
              555                      560                      565

gtc gaa gag gga tac gga gat gac atc ccg aca cta cta gaa aca ctt      2023
Val Glu Glu Gly Tyr Gly Asp Asp Ile Pro Thr Leu Leu Glu Thr Leu
              570                      575                      580

gttcctgtat agtaaatagc gctcccagtt cagcgccaa gtggctctgc cgcacgtcgc      2083

tgacggaggc tgagctcacg ccggcgagat atgacacaca ataaaattcc gcctcgcagg      2143

cgaacgcagt tgccgacgga acacagaccc cattcgcgaa ggcagctcgc ccagcacagc      2203

aatgctccgc ttctcgccac tcaagtgcct acagcgtatc aggcgcctct ctacagacgt      2263

cgggtttgtg ttgacatcg acggcgctcct cctccacggc gaaaagccca ttccgggtgc      2323

ggccgaggcg ctcaggctgc tggaccgcca acgcataccc ttcattattgc tactaacgg      2383

cggcgggaaa ctagaggccc agcgcacgac gcagatttcc aagctcctgg atgttgacat      2443

ccggccgcag ca                                                         2455

```

<210> 31

<211> 583

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 125

<400> 31

Met Gly Arg Glu Gly Val Pro Leu Ile Ala Ser Leu Asn Ala Ser Asn
 1 5 10 15

Glu Val Met Leu Leu Val Gly Thr Gly Gly Ser Gly Arg Val Val Ser
 20 25 30

Arg Leu Arg Gly Leu Leu Asn Thr Gly Ala Gln Cys Ile Val Val Gln
 35 40 45

Pro Ser Asn Ser Lys Leu Val Ala Gln Leu Tyr Arg Glu Phe Ala Gly
 50 55 60

Glu Pro Leu Leu Arg Ile Leu Asp Arg Ala Phe Ala Leu Glu Asp Leu
 65 70 75 80

Thr Thr Leu Gly Arg Tyr Glu Ala Ser Asn Val Val Asp Arg Val Phe
 85 90 95

Leu Thr Pro Gly His Ser Ala Asp Glu Ile Leu Gln Ala His Ala Val
 100 105 110

Tyr Arg Gln Cys Ile Lys Leu Arg Ile Pro Ile Asn Thr Ser Gln Arg
 115 120 125

Pro Glu Leu Ser Thr Phe Ser Leu Pro Ser Thr Tyr Arg Asp Pro Lys
 130 135 140

Gly Ser Gly Leu Gln Val Ala Val Thr Thr Asp Gly Arg Gly Cys Val
 145 150 155 160

Leu Ala Asn Arg Ile Arg Arg Glu Ile Val Ser Lys Leu Pro Ala Asn
 165 170 175

Ile Ser Glu Val Val Ser Asn Met Gly Arg Leu Arg Asp Cys Ile Thr
 180 185 190

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

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

Arg Phe Asn Lys Leu Val Gln Glu Phe Arg Met Lys Glu Ser Asp Gln
 225 230 235 240

Arg Leu Lys Arg Leu Arg Trp Leu Ser Gln Ile Met Glu Tyr Tyr Pro
 245 250 255
 Leu Ser Glu Leu Ala Asp Val Ser Val Asp Lys Leu Leu Asp Glu Gln
 260 265 270
 Thr Gly Ala Ala Cys Ser Gly Thr Thr Ala Thr Thr Gly Ala Ala Ser
 275 280 285
 Val Thr Gly Ala Thr Glu Asp Thr Ala Ala Glu Pro Ala Glu Gly Val
 290 295 300
 Ala Ser Asp Ser Ser Gly Ala Arg Asp Leu Ala Arg Ser Ala Ala Gly
 305 310 315 320
 Ser Ile Ala Leu Val Gly Ser Gly Pro Gly Ser Leu Ser Met Leu Thr
 325 330 335
 Leu Gly Ala Leu Ser Glu Ile Lys Ser Ala Asp Leu Val Leu Ala Asp
 340 345 350
 Lys Leu Val Pro Gln Glu Val Met Asn Thr Ile Pro Ala Gly Thr Glu
 355 360 365
 Thr Phe Ile Ala Arg Lys Phe Pro Gly Asn Ala Glu Arg Ala Gln Glu
 370 375 380
 Glu Leu Ala Glu Lys Ala Leu Ala Ala Leu Arg Glu Gly Lys Arg Val
 385 390 395 400
 Val Arg Leu Lys Gln Gly Asp Pro Tyr Ile Phe Gly Arg Gly Gly Glu
 405 410 415
 Glu Tyr Leu Leu Phe Lys Arg His Gly Phe Glu Pro Arg Val Val Pro
 420 425 430
 Gly Ile Ser Ser Ala Leu Ala Ala Thr Thr Val Ala His Ile Pro Ala
 435 440 445
 Thr Gln Arg Asp Val Ala Glu Gln Val Leu Ile Cys Thr Gly Thr Gly
 450 455 460
 Arg Arg Gly Val Leu Pro Pro Ala Pro Glu Phe Val Ala Ala Arg Thr
 465 470 475 480
 Thr Ile Phe Leu Met Ala Leu His Arg Val Gly Asp Leu Val Gln Ala
 485 490 495

Leu Leu Glu Gln Gln Trp Asp Pro Gln Val Pro Ala Ala Ile Val Glu
 500 505 510

Arg Ala Ser Cys Pro Asp Gln Arg Val Thr Arg Thr Thr Leu Ala Ser
 515 520 525

Val Pro Glu Val Val Glu Gln Ile Gly Ser Arg Pro Pro Gly Leu Leu
 530 535 540

Ile Val Gly His Ala Val Thr Ser Leu Lys Gly Leu Gly Ser Glu Gln
 545 550 555 560

Ala Gly Ala Leu His Tyr Ser Val Glu Glu Gly Tyr Gly Asp Asp Ile
 565 570 575

Pro Thr Leu Leu Glu Thr Leu
 580

<210> 32

<211> 447

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 107

<400> 32
 gccgcgatat aattaagaaa cggttggaat cggcagaact agaggttgaa gagataatca 60
 cagtatcagt accagccac ctccatgata ttagtgagca gcctttgaac gtgtgttcac 120
 cgataactct tgggtcgtgc tatttagtcc tcagggcacc aatgaactgc ttcccctact 180
 gagacagggg gctgccaaga ttgcagctat tggaccaaca actgcgcaat atctacttga 240
 caatgccact ccaccaatat taacgagccc taaacctgag ccctcgagct tgtacaatgc 300
 aataaaagat tatttggcag cgtctcctat gggcaattag ttggagggca ctatgtatgt 360
 atagctttat acttatgaaa tgggtgcagaa cacgcatgtg ttcacgttac cagtttatgg 420
 agtctatgat caaataacaa atatattt 447

<210> 33

<211> 69

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 107

<400> 33

Glu Arg Val Phe Thr Asp Asn Ser Trp Val Val Leu Phe Ser Pro Gln
1 5 10 15

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

Ala Ala Ile Gly Pro Thr Thr Ala Gln Tyr Leu Leu Asp Asn Ala Thr
35 40 45

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

Ala Ile Lys Asp Tyr
65

<210> 34

<211> 1626

<212> DNA

<213> Ashbya gossypii

<220>

<221> CDS

<222> (250)..(1035)

<223>

<220>

<221> misc_feature

<223> Oligo 107

<400> 34

tcacccttgg gatccgctga ctggctactg taattctatg tagtggatta gtacgataag	60
ttacttttgt atgatagatg taatcacatc tggctattaa aatgactcag ccgaggtaaa	120
tctaacgtcc cttcacaagg gtgttcctgt gtggacttcc gcctgaattt ttatagatat	180
atagatactc tactcatgaa caacctgcaa ccgaataagc attagtgccca ggagaagaga	240
accgtggaa atg ggg caa gta gaa aaa atc ata ttc ctt aag aat aag aca	291
Met Gly Gln Val Glu Lys Ile Ile Phe Leu Lys Asn Lys Thr	
1 5 10	
gta cca gag gac cat tac gag acg att ttt gaa tcg aat ggc ttc cag	339
Val Pro Glu Asp His Tyr Glu Thr Ile Phe Glu Ser Asn Gly Phe Gln	
15 20 25 30	
act cac ttt gta ccc ata ata acc cat gaa cac ctg cca gat gag gtt	387
Thr His Phe Val Pro Ile Ile Thr His Glu His Leu Pro Asp Glu Val	
35 40 45	
cgc ggt cga cta tcc gac gcg aat tac atg aaa agg ttg aat tgt ttg	435
Arg Gly Arg Leu Ser Asp Ala Asn Tyr Met Lys Arg Leu Asn Cys Leu	
50 55 60	
gtg gta acc tct cag agg act gtg gag tgt ctc tat gag gac gtt ctg	483
Val Val Thr Ser Gln Arg Thr Val Glu Cys Leu Tyr Glu Asp Val Leu	
65 70 75	
ccc tct ctt cca gct gaa gca cgc aaa tct ctt ctc aat acg cca gta	531
Pro Ser Leu Pro Ala Glu Ala Arg Lys Ser Leu Leu Asn Thr Pro Val	
80 85 90	
ttc gtg gtt ggg cgt gcc act cag gaa ttt atg gag aga tgc ggc ttt	579
Phe Val Val Gly Arg Ala Thr Gln Glu Phe Met Glu Arg Cys Gly Phe	
95 100 105 110	
acg gac gtg aga ggg gga tct gag act ggt aat ggc gtt ttg cta gcg	627
Thr Asp Val Arg Gly Gly Ser Glu Thr Gly Asn Gly Val Leu Leu Ala	
115 120 125	
gag tta atg tta aat atg atc cag aag ggc gat atc agc tcc tcg gga	675
Glu Leu Met Leu Asn Met Ile Gln Lys Gly Asp Ile Ser Ser Ser Gly	
130 135 140	
act ttt ctg ctg ctc gtg ggc gag atc cgc cgc gat ata att aag aaa	723
Thr Phe Leu Leu Leu Val Gly Glu Ile Arg Arg Asp Ile Ile Lys Lys	
145 150 155	
cgg ttg gaa tcg gca gaa cta gag gtt gaa gag ata atc acg tat cgt	771
Arg Leu Glu Ser Ala Glu Leu Glu Val Glu Glu Ile Ile Thr Tyr Arg	
160 165 170	
acc agc cca ctc cat gat att agt gag cgc ttt gaa cgt gtg ttc acc	819

```

Thr Ser Pro Leu His Asp Ile Ser Glu Arg Phe Glu Arg Val Phe Thr
175                      180                      185                      190

gat aac tct tgg gtc gtg cta ttt agt cct cag ggc acc aat gaa ctg      867
Asp Asn Ser Trp Val Val Leu Phe Ser Pro Gln Gly Thr Asn Glu Leu
                      195                      200                      205

ctt ccc cta ctg aga cag ggg gct gcc aag att gca gct att gga cca      915
Leu Pro Leu Leu Arg Gln Gly Ala Ala Lys Ile Ala Ala Ile Gly Pro
                      210                      215                      220

aca act gcg caa tat cta ctt gac aat gcc act cca cca ata tta acg      963
Thr Thr Ala Gln Tyr Leu Leu Asp Asn Ala Thr Pro Pro Ile Leu Thr
                      225                      230                      235

agc cct aaa cct gag ccc tcg agc ttg tac aat gca ata aaa gat tat      1011
Ser Pro Lys Pro Glu Pro Ser Ser Leu Tyr Asn Ala Ile Lys Asp Tyr
                      240                      245                      250

ttg gca gcg tct cct atg ggc aat tagttggagg gcactatgta tgtatagctt      1065
Leu Ala Ala Ser Pro Met Gly Asn
255                      260

tatacttatg aaatggtgca gaacacgcat gtgttcacgt taccagttta tggagtctat      1125

gatcaaataa caaatatctt atcgcgctgg tatatgtgca caacttccac cactttgtcc      1185

ttctcgttgc tatagatctc tgcctttgga tatattaaat agccgcgccg caaagagatc      1245

aactgaatcg agaatactaa atggtcgcct actattacag aacggcgcct gccgttaatt      1305

gaccagtcag ggctgttgct gaactcgaca atatatgtca atggttccct ttccatcacc      1365

cactgcgtag acaagttgca cacttctaca tcaattgtga ccgcatcccc cacatgtaat      1425

acttcgtggc ttgtggcaga caagcgaaca ctgaatagct gtttgatgga tggaagatgc      1485

actgggacaa ttaacgtatt gggttggata ctaaaccgtg tttgtcggga gagaactcta      1545

tgccagaagc taataacttg agatagctac caaaccgggt caagatctca gaatcactaa      1605

tgggacgagg agaggtcatt c                                          1626

<210> 35
<211> 262
<212> PRT
<213> Ashbya gossypii

<220>
<221> misc_feature
<223> Oligo 107
<400> 35

```

Met Gly Gln Val Glu Lys Ile Ile Phe Leu Lys Asn Lys Thr Val Pro
 1 5 10 15
 Glu Asp His Tyr Glu Thr Ile Phe Glu Ser Asn Gly Phe Gln Thr His
 20 25 30
 Phe Val Pro Ile Ile Thr His Glu His Leu Pro Asp Glu Val Arg Gly
 35 40 45
 Arg Leu Ser Asp Ala Asn Tyr Met Lys Arg Leu Asn Cys Leu Val Val
 50 55 60
 Thr Ser Gln Arg Thr Val Glu Cys Leu Tyr Glu Asp Val Leu Pro Ser
 65 70 75 80
 Leu Pro Ala Glu Ala Arg Lys Ser Leu Leu Asn Thr Pro Val Phe Val
 85 90 95
 Val Gly Arg Ala Thr Gln Glu Phe Met Glu Arg Cys Gly Phe Thr Asp
 100 105 110
 Val Arg Gly Gly Ser Glu Thr Gly Asn Gly Val Leu Leu Ala Glu Leu
 115 120 125
 Met Leu Asn Met Ile Gln Lys Gly Asp Ile Ser Ser Ser Gly Thr Phe
 130 135 140
 Leu Leu Leu Val Gly Glu Ile Arg Arg Asp Ile Ile Lys Lys Arg Leu
 145 150 155 160
 Glu Ser Ala Glu Leu Glu Val Glu Glu Ile Ile Thr Tyr Arg Thr Ser
 165 170 175
 Pro Leu His Asp Ile Ser Glu Arg Phe Glu Arg Val Phe Thr Asp Asn
 180 185 190
 Ser Trp Val Val Leu Phe Ser Pro Gln Gly Thr Asn Glu Leu Leu Pro
 195 200 205
 Leu Leu Arg Gln Gly Ala Ala Lys Ile Ala Ala Ile Gly Pro Thr Thr
 210 215 220
 Ala Gln Tyr Leu Leu Asp Asn Ala Thr Pro Pro Ile Leu Thr Ser Pro
 225 230 235 240
 Lys Pro Glu Pro Ser Ser Leu Tyr Asn Ala Ile Lys Asp Tyr Leu Ala
 245 250 255

Ala Ser Pro Met Gly Asn
260

<210> 36

<211> 368

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 136

<400> 36
gatctcccat gtgtccaccg gtggtggtgc ctccctagag ctattggagg gtaaggacct 60
accaggtgtc accttcttgt ccagcaagca gtagatgctg tccgcgcgcc tgccggcctt 120
gaatagataa taaactatat acgcacgtaa tggaaagttg agaaactact atatgcgctt 180
gttgcgctgc tcgtccgtgg gctccagccg cgggaacccg agcagctcca tgatgcggta 240
ctcgtccgcg ctctcgacgc acgcgccgct cgacttgtca aacaggccgt gctgattcag 300
cgtcatgccc tgcgcggagg ccccgcgctc gcagagcctt gttgaaaagt gtogttgccg 360
atgaaagt 368

<210> 37

<211> 30

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 136

<400> 37

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

Gly Lys Asp Leu Pro Gly Val Thr Phe Leu Ser Ser Lys Gln
 20 25 30

<210> 38

<211> 2000

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (447) .. (1790)

<223>

<220>

<221> misc_feature

<223> Oligo 136

<400> 38

ctgcagcacg tggtcctgcc gggtttcgag ttcgccgcac agctcccgca ccgcatgta 60

ccgaagatct acatcctcct gctcgcgata gcgcatacag caccgcagcg cgctctccat 120

ctccgtctcc atccttgccg ttgcccgtat ctgccaacaa gccgtccaga acgccgctgc 180

ttagtacgag ctcgtaacgt gcgactgcca gcttcctggt ctgtggaagc cggcttctcc 240

cctggctttt tccgtagcat gacgcggggg ctcaccagag aacgaggaag aatttccagt 300

tggcgagcac aggccaggcg ggctccctac gggtagcgtg gggcgagcaa ggcagggcaa 360

tgatgcaata agcgcaggcc ggcacgcaat tgagcgagca acgtgcacga ggatttttct 420

gggggagaaa aaagcgagta taaagg atg ggc cgg tcg tgg ggt ctg ctc tgg 473
 Met Gly Arg Ser Trp Gly Leu Leu Trp

1

5

atg gcg tgc ttg ctg tac ggc gtg ttc aca gag tgt cgc aga tct tca 521

Met Ala Cys Leu Leu Tyr Gly Val Phe Thr Glu Cys Arg Arg Ser Ser
 10 15 20 25

cag cac gca cac aca gga aaa atg tcg cta tcc tct aag cta aca gtc 569

Gln His Ala His Thr Gly Lys Met Ser Leu Ser Ser Lys Leu Thr Val
 30 35 40

aag gac ctg agc ctt gcg ggc aaa cgc gtg ttc atc cgc gtg gac ttc Lys Asp Leu Ser Leu Ala Gly Lys Arg Val Phe Ile Arg Val Asp Phe 45 50 55	617
aac gtg ccg cta gac ggg aag acg atc acg tcg aac cag cgc att gtg Asn Val Pro Leu Asp Gly Lys Thr Ile Thr Ser Asn Gln Arg Ile Val 60 65 70	665
gcc gcg cta ccc acg atc aag tac gtg ttg gag cag ggc ccc aag gcg Ala Ala Leu Pro Thr Ile Lys Tyr Val Leu Glu Gln Gly Pro Lys Ala 75 80 85	713
gtg gtg ttg gcg tcg cac cta ggg cgg ccc aac ggc gag cgc aac gag Val Val Leu Ala Ser His Leu Gly Arg Pro Asn Gly Glu Arg Asn Glu 90 95 100 105	761
aag tac tcg ttg gcg ccg gtg gcc gcg gag ctg gag aag ctg ctg ggc Lys Tyr Ser Leu Ala Pro Val Ala Ala Glu Leu Glu Lys Leu Leu Gly 110 115 120	809
cag aag gtg aac ttc ctg gac gac tgc gtg ggc gag cac gtg acc gct Gln Lys Val Asn Phe Leu Asp Asp Cys Val Gly Glu His Val Thr Ala 125 130 135	857
gcg gtc aac ggc gcg gct gcg ggg tcc gtg ttc ctg ctt gag aac ctg Ala Val Asn Gly Ala Ala Ala Gly Ser Val Phe Leu Leu Glu Asn Leu 140 145 150	905
cgg ttc cac atc gag gag gag ggc tcg cgc aag gtg gac ggc gag aag Arg Phe His Ile Glu Glu Glu Gly Ser Arg Lys Val Asp Gly Glu Lys 155 160 165	953
gtc aag gcg tct gcg gag gac gtg cag aag ttc cgc cag ggc ctc atg Val Lys Ala Ser Ala Glu Asp Val Gln Lys Phe Arg Gln Gly Leu Met 170 175 180 185	1001
tcg ctc gcg gac gtc tac gtc aac gac gcc ttc ggc acc gcg cac aga Ser Leu Ala Asp Val Tyr Val Asn Asp Ala Phe Gly Thr Ala His Arg 190 195 200	1049
gcg cac tcc tcg atg gtc ggc ttc gag ctg cct gag agg gcc ggc ggc Ala His Ser Ser Met Val Gly Phe Glu Leu Pro Glu Arg Ala Gly Gly 205 210 215	1097
ttc ttg ttg tcg cgc gag ctg gag tac ttc agc aag gcg ttg gag aac Phe Leu Leu Ser Arg Glu Leu Glu Tyr Phe Ser Lys Ala Leu Glu Asn 220 225 230	1145
cct acg aga ccc ttc ttg gcg atc ctg ggc ggt gcc aag gtt gcc gac Pro Thr Arg Pro Phe Leu Ala Ile Leu Gly Gly Ala Lys Val Ala Asp 235 240 245	1193
aag atc cag ttg atc gac aac ttg ttg gac aag gtc gac tcg att gtg Lys Ile Gln Leu Ile Asp Asn Leu Leu Asp Lys Val Asp Ser Ile Val 250 255 260 265	1241
atc ggc ggt ggc atg gcc ttc acc ttc aag aag gtg ctt gag aac atg Ile Gly Gly Gly Met Ala Phe Thr Phe Lys Lys Val Leu Glu Asn Met 270 275 280	1289
gag atc ggt aac tcc atc tac gac aag gcg ggt gcc gag atc gtt cct Glu Ile Gly Asn Ser Ile Tyr Asp Lys Ala Gly Ala Glu Ile Val Pro	1337

285	290	295	
aag cta gcg gag aag gcc aag aag aac ggc gtc aag atc gtg ttg cct Lys Leu Ala Glu Lys Ala Lys Lys Asn Gly Val Lys Ile Val Leu Pro 300 305 310			1385
gtg gac ttc gtg atc ggc gac gac ttc tcc cag gac gcc aac acc aag Val Asp Phe Val Ile Gly Asp Asp Phe Ser Gln Asp Ala Asn Thr Lys 315 320 325			1433
att gtc tct gcc tcc gag ggc att cca tcc ggc tgg gag ggt cta gac Ile Val Ser Ala Ser Glu Gly Ile Pro Ser Gly Trp Glu Gly Leu Asp 330 335 340 345			1481
tgt ggt cct gag tcc aga aag ctg ttc agc gag acc atc gcg agc gcg Cys Gly Pro Glu Ser Arg Lys Leu Phe Ser Glu Thr Ile Ala Ser Ala 350 355 360			1529
aag acc atc gtg tgg aac ggc cca cca ggt gtc ttt gag att cca aag Lys Thr Ile Val Trp Asn Gly Pro Pro Gly Val Phe Glu Ile Pro Lys 365 370 375			1577
ttc tcc gag ggt acc cag gcc atg ctt gcg gcc gct gtc aag gcc tcc Phe Ser Glu Gly Thr Gln Ala Met Leu Ala Ala Val Lys Ala Ser 380 385 390			1625
gag gct ggc agc acc gtc atc atc ggc ggt ggt gac acc gct acc gtg Glu Ala Gly Ser Thr Val Ile Ile Gly Gly Gly Asp Thr Ala Thr Val 395 400 405			1673
gcc aag aag tac ggt gtg gtc gag aag atc tcc cat gtg tcc acc ggt Ala Lys Lys Tyr Gly Val Val Glu Lys Ile Ser His Val Ser Thr Gly 410 415 420 425			1721
ggt ggt gcc tcc cta gag cta ttg gag ggt aag gac cta cca ggt gtc Gly Gly Ala Ser Leu Glu Leu Leu Glu Gly Lys Asp Leu Pro Gly Val 430 435 440			1769
acc ttc ttg tcc agc aag cag tagatgcgtc cgcgcgcctg ccggcttgaa Thr Phe Leu Ser Ser Lys Gln 445			1820
tagataataa actatatacg caacgtaatg gaaagttgag aaactactat atgcgcctgt			1880
tgcgctgctc gtccgtgggc tccagccgcg ggaacccgag cagctccatg atgcagtact			1940
cgtccggggct ctcgacgcac gcgccgctcg acttgtcaaa caggccgtgc tgattcagcg			2000

<210> 39

<211> 448

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 136

<400> 39

Met Gly Arg Ser Trp Gly Leu Leu Trp Met Ala Cys Leu Leu Tyr Gly
1 5 10 15

Val Phe Thr Glu Cys Arg Arg Ser Ser Gln His Ala His Thr Gly Lys
20 25 30

Met Ser Leu Ser Ser Lys Leu Thr Val Lys Asp Leu Ser Leu Ala Gly
35 40 45

Lys Arg Val Phe Ile Arg Val Asp Phe Asn Val Pro Leu Asp Gly Lys
50 55 60

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

Tyr Val Leu Glu Gln Gly Pro Lys Ala Val Val Leu Ala Ser His Leu
85 90 95

Gly Arg Pro Asn Gly Glu Arg Asn Glu Lys Tyr Ser Leu Ala Pro Val
100 105 110

Ala Ala Glu Leu Glu Lys Leu Leu Gly Gln Lys Val Asn Phe Leu Asp
115 120 125

Asp Cys Val Gly Glu His Val Thr Ala Ala Val Asn Gly Ala Ala Ala
130 135 140

Gly Ser Val Phe Leu Leu Glu Asn Leu Arg Phe His Ile Glu Glu Glu
145 150 155 160

Gly Ser Arg Lys Val Asp Gly Glu Lys Val Lys Ala Ser Ala Glu Asp
165 170 175

Val Gln Lys Phe Arg Gln Gly Leu Met Ser Leu Ala Asp Val Tyr Val
180 185 190

Asn Asp Ala Phe Gly Thr Ala His Arg Ala His Ser Ser Met Val Gly
195 200 205

Phe Glu Leu Pro Glu Arg Ala Gly Gly Phe Leu Leu Ser Arg Glu Leu
210 215 220

Glu Tyr Phe Ser Lys Ala Leu Glu Asn Pro Thr Arg Pro Phe Leu Ala
 225 230 235 240

Ile Leu Gly Gly Ala Lys Val Ala Asp Lys Ile Gln Leu Ile Asp Asn
 245 250 255

Leu Leu Asp Lys Val Asp Ser Ile Val Ile Gly Gly Gly Met Ala Phe
 260 265 270

Thr Phe Lys Lys Val Leu Glu Asn Met Glu Ile Gly Asn Ser Ile Tyr
 275 280 285

Asp Lys Ala Gly Ala Glu Ile Val Pro Lys Leu Ala Glu Lys Ala Lys
 290 295 300

Lys Asn Gly Val Lys Ile Val Leu Pro Val Asp Phe Val Ile Gly Asp
 305 310 315 320

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

Ile Pro Ser Gly Trp Glu Gly Leu Asp Cys Gly Pro Glu Ser Arg Lys
 340 345 350

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

Pro Pro Gly Val Phe Glu Ile Pro Lys Phe Ser Glu Gly Thr Gln Ala
 370 375 380

Met Leu Ala Ala Ala Val Lys Ala Ser Glu Ala Gly Ser Thr Val Ile
 385 390 395 400

Ile Gly Gly Gly Asp Thr Ala Thr Val Ala Lys Lys Tyr Gly Val Val
 405 410 415

Glu Lys Ile Ser His Val Ser Thr Gly Gly Gly Ala Ser Leu Glu Leu
 420 425 430

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

<210> 40

<211> 1006

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<222> (361)..(361)

<223> n = unknown nucleotide

<220>

<221> misc_feature

<222> (388)..(388)

<223> n = unknown nucleotide

<220>

<221> misc_feature

<223> Oligo 157

<400> 40
gatcgcgccg caccacctcc tcccacggcg gccgcttctc cggcgcaaca ggaggcttgt 60
ggcggtcgta caggcacctg gtagtactgc accttcacac gcgtgcctgt cgatccgctt 120
gcgctgccag gtcatagaac ttgggcaaatt tgaggctccg gtagtcccct ccgttgctga 180
taaattggcg aagccggtct tcgtagatgc cccttgacgg gcttgaactg cgggtcccag 240
ttgaccgata tcgctgctca tagcgctgct tccttcctgc cttgggtccgt agtctcgctt 300
gccgctcgca ccttttaata tccttaccgg cgcgggtttc gtaatgcagt tttcscctgc 360
ngtycgscgg cagtcacccg aaaccctncc cggaaaaacg cttttccccg caggggcgcg 420
catgaaacct gctgctgccg gcctgttgct tgcctccggg gcacgttgat tattcagcta 480
tactgattat atatactcta gttgtccaaa atgtgcatct cctggtcctt ctgatgtgc 540
tggatcaagt cgctgtgctg cgacttaatc tgctccaagg ccgtgacctc gggcaacttc 600
acaactgaggc cgttgatgag cgagtgtctg tgcaggatac tcccgccgat ctgctcgatg 660
gactctctga ccttggcgag gtcggccccc ttcttcaacg tgatgatata agatgattca 720
gacatgcttg gtgctgggac tgatggtagt catctaccgg cgttcgctg cccttttata 780
tacggccgag cccggcaagg ggcccctagg ccccgtttc gcttgccgag gccactgttg 840
ctggctactt acttaggctt actgtacyta gattgctctg ggctgtgcgg tcagaggtgc 900

tgtaatcagg attgtcgagc gcacgcccc gctcacgtga cagaacggcc atcaggcggg 960
aacgcgcgcg tagcgagacc cactcgccgt caatcagagc cactag 1006

<210> 41

<211> 67

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 157

<400> 41

Ser	Tyr	Ile	Ile	Thr	Leu	Lys	Lys	Gly	Ala	Asp	Leu	Ala	Lys	Val	Arg
1				5				10					15		

Glu	Ser	Ile	Glu	Gln	Ile	Gly	Gly	Ser	Ile	Leu	His	Glu	His	Ser	Leu
		20					25					30			

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

Ile	Lys	Ser	Gln	His	Asp	Asp	Leu	Ile	Gln	His	Ile	Glu	Lys	Asp	Gln
	50				55				60						

Glu Met His
65

<210> 42

<211> 1658

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (809)..(1033)

<223>

<220>

<221> misc_feature

<223> Oligo 157

<400> 42

```

gttatgatga gccccgccag ccccgcgggga aatagagcag aatgtacatg ccgtgttttc      60
gtccggacgt cctttcgtgc agttgtgtgc gactttttca gactttccgc ggagccggca      120
gcgcaacctc ttccccact atcgcaatga tccgccatac cgtccccgtt gcgcgccgta      180
tcctgtctcag taggccttct gcactcctgg cgcgccccgc cgtacgcata gccgcgaggc      240
gggccgcgtg gcgcgccatg tcctccttct ccggcggcat gcgcgctccc ccggagctgg      300
cccagctcac gccgcaggcc taccacaagc aggccgacga gttcctgaac gacctgctgg      360
accgtctcga agcgtctgggc gacgagcgcc ccgacatcat cgcagactcc gagtactccc      420
agggtgtgct cacgtgttcc gtacctgtc tgggcaccta tgtcattaac aaacagcccc      480
caaacaagca gatatggctc tcgtcgccag tatccggccc caaccgctag tggctctgat      540
tgacggcgag tgggtctcgc tacgcgacgg ttcccgctg atggccgttc tgtcacgtga      600
gctggggcgt gcgctcgaca atcctgatta cagcacctct gaccgcacag cccagagcaa      660
tctargtaca gtagcaccta agtaagtagc cagcaacagt ggctgcgca agcgaaacgg      720
gggcctaggg gccccttgcc gggctcggcc gtatataaaa gggcacgcga acgcgggtag      780
atgactacca tcagtcccag caccaagc atg tct gaa tca tct tat atc atc      832
                        Met Ser Glu Ser Ser Tyr Ile Ile
                        1                      5

acg ttg aag aag ggg gcc gac ctc gcc aag gtc aga gag tcc atc gag      880
Thr Leu Lys Lys Gly Ala Asp Leu Ala Lys Val Arg Glu Ser Ile Glu
  10                      15                      20

cag atc ggc ggg agt atc ctg cac gag cac tcg ctc atc aac ggc ctc      928
Gln Ile Gly Gly Ser Ile Leu His Glu His Ser Leu Ile Asn Gly Leu
  25                      30                      35                      40

agt gtg aag ttg ccc gag gtc acg gcc ttg gag cag att aag tcg cag      976
Ser Val Lys Leu Pro Glu Val Thr Ala Leu Glu Gln Ile Lys Ser Gln
                45                      50                      55

cac gac gac ttg atc cag cac atc gag aag gac cag gag atg cac att      1024
His Asp Asp Leu Ile Gln His Ile Glu Lys Asp Gln Glu Met His Ile
                60                      65                      70

ttg gac aac tagagtatat ataatacagta tagctgaata atcaacgtgc      1073
Leu Asp Asn

```

75

```

cccgaggca gacaacaggc cggcagcagc aggtttcatg gcgcgcctg cggggaagcg 1133
ttttccggga ggtttcggtg actgcggccg acgcagcgaa actgcatacg aaaccgcgcg 1193
gtaggatata aaggtgcgag cggcaggcga gactacggac caaggcagga aggagcagcg 1253
ctatgagcag cgatatcgtc aactgggacc cgcagttcaa gcccgtaag ggcattctacg 1313
aagaccggct tcgccaattt atcgacaacg gaggggacta ccgggacctc aatttgccca 1373
agttctatga ccggcagcgc aagcggatcg acagcacgcg tgtgaagggtg cagtactacc 1433
aggtgcctgt acgaccgcca caagcctcct gttgcgccgg agaagcggcc gccgtgggag 1493
gaggtggtgc ggcgcgatcg tgctggagaa atcgaatgga agatgtgtac ctcgaccccg 1553
tcgggccgag ctggtccact acgaggttct aggtgagttg gacatcccg aatcgtggag 1613
cagctctggg gaacagataa tattctagtg ggactgcgag aaaa 1658

```

<210> 43

<211> 75

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 157

<400> 43

```

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

```

```

Ala Lys Val Arg Glu Ser Ile Glu Gln Ile Gly Gly Ser Ile Leu His
20           25           30

```

```

Glu His Ser Leu Ile Asn Gly Leu Ser Val Lys Leu Pro Glu Val Thr
35           40           45

```

```

Ala Leu Glu Gln Ile Lys Ser Gln His Asp Asp Leu Ile Gln His Ile
50           55           60

```

```

Glu Lys Asp Gln Glu Met His Ile Leu Asp Asn
65           70           75

```

<210> 44

<211> 1599

<212> DNA

<213> Ashbya gossypii

<220>

<221> misc_feature

<222> (331)..(331)

<223> n = unknown nucleotide

<220>

<221> misc_feature

<222> (343)..(343)

<223> n = unknown nucleotide

<220>

<221> misc_feature

<223> Oligo 108

<400> 44

gatcgatata cagggttttgg ggcatttgca actcctcttg cgggatatgt caccgctgcg	60
gcatccgaac agggggccctg aattactggc ggttgaccca gatctgttca gccgcatggt	120
cttcacgaag ccgtctcagg attcgctggt cctgaagagc aagggaacct tgccgtcgcg	180
gattottccc catttgtacc ttggttcact cgagcacgcg catgacccaa acctgctccg	240
cagattgggt atcagaaata ttgtgtcagt cggggagact ttatcatggc tacaatcccc	300
tcactcccgg gaggttgcta aaacagaaga nagcaaaca ganaccatca cggacgcgta	360
tgtgaccgca cctgctgaaa ccgccgcggc aacgtccacg accgcatcca ctggtagtct	420
taatttggtg aagcgaaaca gatcaccag cgcgctgaa attaaaccgt taaaatcctc	480
atcctgtaaa actacagtct tctcgaagga tggattcgaa atatgccaca caagcaactt	540
gggagacaat ggtgcagatc cgatgctcac tcagctggac caaatgctct ctttcatcaa	600
tcggtgttat gaggccggcg aaaaggtact ggttcactgc atggtcggcg tttctcgcag	660

```

•   tgcgacagta tgcatagccg aatgcatgag acgtctaaac tgcgcagtga tccaggccta   720
    cctgtttgtc cgtgtgcgcc gcttaaacad tataatccag cccaatctca tgttcaccta   780
    cgagctcctc aagtggcagg agctcaagca ccctgacgtt ctcaagattg actggcatat   840
    catgtgtagg gccatcaacg aactcaacag taattatata aaggacatga tttgacctac   900
    cagttgagct tggttgcccc ctctcacatt ttgtactatt tagaggattt tacgatacca   960
    taaactactc tagtgtaaca tatttcttca acgagtccgg aatctcatcg tagagcttct  1020
    tttcctgcaa ccacttcttc gagaatacgc ggtcggcata cttgtgtcca gtgtcacaga  1080
    ggatcgtgac gacattgctt cccgggtgga gttccctggc cacacgcaca gcgccagcga  1140
    cgttcagccc cgaagtgcc a cgaggaata gaccttcctc gtccagaagc ctgtaggtca  1200
    tgttcaggga ctctgggtcc ggaacaaaca cggcgtcgtc aatgatata acgctggctt  1260
    cgaggttctt tgtaatccgc ccttggtcca cgccctcagt gaacgagctg ccctccctct  1320
    ccaagttgcc ggtcttgacg aatgaataga cgaccgagcc tggcgggtcc gtgaccacca  1380
    gtttgcatth gcgctccgtg gcctcggcca ggtaccgcgc gatcccgag aacgttcctc  1440
    ctggtacctg tggagcaagg tgaatgcac cacctcgagc ccagcgttgc gcatttgctg  1500
    caagatctcg ggaccggtcg tcgcatagtg cgccttccaa ttggcctcgt tgtcaaactg  1560
    gttggtccag acagcattgt ccaatgccgc ggcgtgatc   1599

```

<210> 45

<211> 46

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 108

<400> 45

```

His Ala Ala Ala Leu Asp Asn Ala Val Trp Thr Asn Gln Phe Asp Asn
1           5           10           15

```

```

Glu Ala Asn Trp Lys Ala His Tyr Ala Thr Thr Gly Pro Glu Ile Leu
20           25           30

```

```

Gln Gln Met Arg Asn Ala Gly Leu Glu Val Asp Ala Phe Thr

```

35

40

45

<210> 46

<211> 157

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 108

<400> 46

Gly Gly Thr Phe Ser Gly Ile Ala Arg Tyr Leu Ala Glu Ala Thr Glu
1 5 10 15

Arg Lys Cys Lys Leu Val Val Thr Asp Pro Pro Gly Ser Val Val Tyr
20 25 30

Ser Phe Val Lys Thr Gly Asn Leu Glu Arg Glu Gly Ser Ser Phe Thr
35 40 45

Glu Gly Val Gly Gln Gly Arg Ile Thr Lys Asn Leu Glu Ala Ser Val
50 55 60

Asp Ile Ile Asp Asp Ala Val Phe Val Pro Asp Pro Glu Ser Leu Asn
65 70 75 80

Met Thr Tyr Arg Leu Leu Asp Glu Glu Gly Leu Phe Leu Gly Gly Thr
85 90 95

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

Pro Gly Ser Asn Val Val Thr Ile Leu Cys Asp Thr Gly His Lys Tyr
115 120 125

Ala Asp Arg Val Phe Ser Lys Lys Trp Leu Gln Glu Lys Lys Leu Tyr
130 135 140

Asp Glu Ile Pro Asp Ser Leu Lys Lys Tyr Val Thr Leu
145 150 155

<210> 47

<211> 2808

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (450)..(1517)

<223>

<220>

<221> misc_feature

<223> Oligo 108

<400> 47

```

ccgctgacgg tgcgcaacct ggtcaagtgg cgcacggcaa ggagctggag cagttcctgg      60
cggacgtggc ggacgtgcgg cgggtgttgc cgagcatcag ggcgccgcgg actcgccgga.    120
agaggcggag cgctttgtgg gcggcctgcg cagcgccctt gaggcgccgg cggcggcgcc      180
tgcgcccgca gcgcgcaaga gatacgagtt cgctccccgg cccgcgcgca actgaaccgc      240
gcacctgtat ataccgcgac atagtgaata aagaatgaga gtcccacagt gcgcctgcgc      300
gcggagcgcg cggcatggtc gccgaccgct gttgtgccgc cgcacgtgat tagtgcacta      360
tggctgtgga tctgtgctga tgtcgggtca ccgtccagtc gctgggtata tcagtacagc      420
agtaggcagc ccggccggaa cacgttacg atg cac gcc cag ctc att agg aag      473
                               Met His Ala Gln Leu Ile Arg Lys
                               1                               5

atg tca gcg ccc aca ctg gcc gcc gga ggc ttc ccg gat gcc att ggc      521
Met Ser Ala Pro Thr Leu Ala Ala Gly Gly Phe Pro Asp Ala Ile Gly
   10                               15                               20

aag acg ccg ctg atc aag ctg cag aag ttg tgc cag gaa ctg ggt cgg      569
Lys Thr Pro Leu Ile Lys Leu Gln Lys Leu Ser Gln Glu Leu Gly Arg
   25                               30                               35                               40

aac atc tac ggg aag gcg gag ttt atg aac ccg ggc gga tcc gtc aag      617
Asn Ile Tyr Gly Lys Ala Glu Phe Met Asn Pro Gly Gly Ser Val Lys
                               45                               50                               55

gac cgt gct gcg ctt tac ctg gtg gaa gaa gcg gaa cgg gcg ggg ctg      665

```

-	Asp	Arg	Ala	Ala	Leu	Tyr	Leu	Val	Glu	Glu	Ala	Glu	Arg	Ala	Gly	Leu	
			60						65					70			
	atc	gac	cct	tcg	cgg	ccg	ggg	acc	ata	gta	gag	ggc	tct	gcc	ggc	aac	713
	Ile	Asp	Pro	Ser	Arg	Pro	Gly	Thr	Ile	Val	Glu	Gly	Ser	Ala	Gly	Asn	
			75					80					85				
	acg	gcg	atc	ggg	ctg	gca	cac	ata	gca	cgc	tcc	agg	ggg	tac	aat	gct	761
	Thr	Ala	Ile	Gly	Leu	Ala	His	Ile	Ala	Arg	Ser	Arg	Gly	Tyr	Asn	Ala	
			90				95					100					
	gtg	ttc	tac	atg	ccg	gac	acg	cag	tcg	cgg	gaa	aag	gtc	aac	ttg	ctg	809
	Val	Phe	Tyr	Met	Pro	Asp	Thr	Gln	Ser	Arg	Glu	Lys	Val	Asn	Leu	Leu	
	105					110					115					120	
	agg	tac	ctg	ggg	acc	gag	gtg	cac	ccg	gta	ccc	gtg	tgc	ccc	agc	aca	857
	Arg	Tyr	Leu	Gly	Thr	Glu	Val	His	Pro	Val	Pro	Val	Cys	Pro	Ser	Thr	
					125					130					135		
	gac	cct	ctg	aac	ttc	aac	aat	cgc	gct	cgg	gat	cac	gcc	gcg	gca	ttg	905
	Asp	Pro	Leu	Asn	Phe	Asn	Asn	Arg	Ala	Arg	Asp	His	Ala	Ala	Ala	Leu	
				140					145					150			
	gac	aat	gct	gtc	tgg	acc	aac	cag	ttt	gac	aac	gag	gcc	aat	tgg	aag	953
	Asp	Asn	Ala	Val	Trp	Thr	Asn	Gln	Phe	Asp	Asn	Glu	Ala	Asn	Trp	Lys	
			155					160					165				
	gcg	cac	tat	gcg	acg	acc	ggt	ccc	gag	atc	ttg	cag	caa	atg	cgc	aac	1001
	Ala	His	Tyr	Ala	Thr	Thr	Gly	Pro	Glu	Ile	Leu	Gln	Gln	Met	Arg	Asn	
			170				175					180					
	gct	ggg	ctc	gag	gtg	gat	gca	ttc	act	tgc	tcc	aca	ggt	acc	gga	gga	1049
	Ala	Gly	Leu	Glu	Val	Asp	Ala	Phe	Thr	Cys	Ser	Thr	Gly	Thr	Gly	Gly	
	185					190					195					200	
	acg	ttc	tcc	ggg	atc	gcg	cgg	tac	ctg	gcc	gag	gcc	acg	gag	cgc	aaa	1097
	Thr	Phe	Ser	Gly	Ile	Ala	Arg	Tyr	Leu	Ala	Glu	Ala	Thr	Glu	Arg	Lys	
					205					210					215		
	tgc	aaa	ctg	gtg	gtc	acg	gac	ccg	cca	ggc	tcg	gtc	gtc	tat	tca	ttc	1145
	Cys	Lys	Leu	Val	Val	Thr	Asp	Pro	Pro	Gly	Ser	Val	Val	Tyr	Ser	Phe	
				220					225					230			
	gtc	aag	acc	ggc	aac	ttg	gag	agg	gag	ggc	agc	tcg	ttc	act	gag	ggc	1193
	Val	Lys	Thr	Gly	Asn	Leu	Glu	Arg	Glu	Gly	Ser	Ser	Phe	Thr	Glu	Gly	
				235				240					245				
	gtg	ggc	caa	ggg	cgg	att	aca	aag	aac	ctc	gaa	gcc	agc	gtt	gat	atc	1241
	Val	Gly	Gln	Gly	Arg	Ile	Thr	Lys	Asn	Leu	Glu	Ala	Ser	Val	Asp	Ile	
				250			255					260					
	att	gac	gac	gcc	gtg	ttt	gtt	ccg	gac	cca	gag	tcc	ctg	aac	atg	acc	1289
	Ile	Asp	Asp	Ala	Val	Phe	Val	Pro	Asp	Pro	Glu	Ser	Leu	Asn	Met	Thr	
	265					270					275					280	
	tac	agg	ctt	ctg	gac	gag	gaa	ggt	cta	ttc	ctc	ggt	ggc	act	tcg	ggg	1337
	Tyr	Arg	Leu	Leu	Asp	Glu	Glu	Gly	Leu	Phe	Leu	Gly	Gly	Thr	Ser	Gly	
					285					290					295		
	ctg	aac	gtc	gct	ggc	gct	gtg	cgt	gtg	gcc	agg	gaa	cta	cca	ccg	gga	1385
	Leu	Asn	Val	Ala	Gly	Ala	Val	Arg	Val	Ala	Arg	Glu	Leu	Pro	Pro	Gly	
				300					305					310			

```

agc aat gtc gtc acg atc ctc tgt gac act gga cac aag tat gcc gac      1433
Ser Asn Val Val Thr Ile Leu Cys Asp Thr Gly His Lys Tyr Ala Asp
      315                      320                      325

cgc gta ttc tcg aag aag tgg ttg cag gaa aag aag ctc tac gat gag      1481
Arg Val Phe Ser Lys Lys Trp Leu Gln Glu Lys Lys Leu Tyr Asp Glu
      330                      335                      340

att ccg gac tcg ttg aag aaa tat gtt aca cta gag tagtttatgg          1527
Ile Pro Asp Ser Leu Lys Lys Tyr Val Thr Leu Glu
      345                      350                      355

tatcgtaaaa tcctctaaat agtacaaaat gtgagagtgg gcaaccaagc tcaactggta  1587

gggtcaaatca tgtcctttat ataattactg ttgagttcgt tgatggccct acacatgata  1647

tgccagtgcaa tcttgagaac gtcaggggtgc ttgagctcct gccacttgag gagctcgtag  1707

atgaacatga gattgggctg gattataatg ttttaagcggc gcacacggac aaacaggtag  1767

gcctggatca ctgcgcagtt tagacgtctc atgcattcgg ctatgcatac tgtcgcactg  1827

cgagaaacgc cgaccatgca gtgaaccagt accttttcgc cggcctcata acaccgattg  1887

atgaaagaga gcatttggtc cagctgagtg agcatcggat ctgcaccatt gtctcccaag  1947

ttgcttgtgt ggcataatttc gaatccatcc ttcgagaaga ctgtagtttt acaggatgag  2007

gattttaacg gtttaatttc aggcgcgctg ggtgatctgt ttcgcttcaa caaattaaga  2067

ctaccagtgg atgcggctgt ggacgttgcc gcggcggttt cagcaggtgc ggtcacatac  2127

gcgtccgtga tggttcttgt ttgcttcttc tgttttagca gcctcccggg agtgagggga  2187

ttgtagccat gataaagtct ccccgactga cacaatatat ctgataccca atctgcggag  2247

caggtttggg tcatgcgcgt gctcgagtga accaaggtag aaatggggaa gaatccgcga  2307

cggcaatggg cccttgctct tcaggaacag cgaatcctga gacggcttcg tgaagaacat  2367

gcggctgaac agatctgggt caaccgccag taattcaggg ccctgttcg gatgccgcag  2427

cggtgacata tcccgcgaaga ggagttgcaa atgccccaaa acctgtatat cgatcggaaa  2487

taggaagtat ggacgctcga cctccgtatg caaccgcaac agaacctcat ccagaggttg  2547

atcccaggca aatatcagat aagccactag caaaaaggaa gtttctgtat aaccatcgga  2607

acagtataat agggaaagct tcttgtgtac cgcgctgaat cggtagatca ggtgacagat  2667

attaacacac gaagtgatgg agtcacagat gaggggtccc aggccaattg agcccagct  2727

cggaaaacgca atgtgtatcg gagcctccat ctgggtcaag tgatgcctga tgtcgtcgtg  2787

gtgtgggaag ggcgtgtgct c                                          2808

```

<210> 48

<211> 356

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 108

<400> 48

Met His Ala Gln Leu Ile Arg Lys Met Ser Ala Pro Thr Leu Ala Ala
 1 5 10 15

Gly Gly Phe Pro Asp Ala Ile Gly Lys Thr Pro Leu Ile Lys Leu Gln
 20 25 30

Lys Leu Ser Gln Glu Leu Gly Arg Asn Ile Tyr Gly Lys Ala Glu Phe
 35 40 45

Met Asn Pro Gly Gly Ser Val Lys Asp Arg Ala Ala Leu Tyr Leu Val
 50 55 60

Glu Glu Ala Glu Arg Ala Gly Leu Ile Asp Pro Ser Arg Pro Gly Thr
 65 70 75 80

Ile Val Glu Gly Ser Ala Gly Asn Thr Ala Ile Gly Leu Ala His Ile
 85 90 95

Ala Arg Ser Arg Gly Tyr Asn Ala Val Phe Tyr Met Pro Asp Thr Gln
 100 105 110

Ser Arg Glu Lys Val Asn Leu Leu Arg Tyr Leu Gly Thr Glu Val His
 115 120 125

Pro Val Pro Val Cys Pro Ser Thr Asp Pro Leu Asn Phe Asn Asn Arg
 130 135 140

Ala Arg Asp His Ala Ala Ala Leu Asp Asn Ala Val Trp Thr Asn Gln
 145 150 155 160

Phe Asp Asn Glu Ala Asn Trp Lys Ala His Tyr Ala Thr Thr Gly Pro
 165 170 175

Glu Ile Leu Gln Gln Met Arg Asn Ala Gly Leu Glu Val Asp Ala Phe
 180 185 190

[illegible]