

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

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
20. Februar 2003 (20.02.2003)

PCT

(10) Internationale Veröffentlichungsnummer
WO 03/014351 A2

(51) Internationale Patentklassifikation⁷: **C12N 15/30**,
15/52, C12Q 1/68, C07K 14/37, C12N 1/15, 15/80, C12P
25/00, C12Q 1/02, C07K 16/14

Wachenheim (DE). **KRÖGER, Burkhard** [DE/DE];
Im Waldhof 1, 67117 Limburgerhof (DE). **REVUELTA
DOVAL, Jose, L.** [ES/ES]; Grillo 11, 4 E, 37001 Sala-
manca (ES).

(21) Internationales Aktenzeichen: PCT/EP02/07696

(74) Anwälte: **KINZEBACH, Werner** usw.; Sternwartstrasse
4, 81679 München (DE).

(22) Internationales Anmeldedatum:
10. Juli 2002 (10.07.2002)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:

101 33 327.7	10. Juli 2001 (10.07.2001)	DE
101 33 328.5	10. Juli 2001 (10.07.2001)	DE
101 33 329.3	10. Juli 2001 (10.07.2001)	DE
101 33 371.4	10. Juli 2001 (10.07.2001)	DE
101 33 372.2	10. Juli 2001 (10.07.2001)	DE
102 09 815.8	6. März 2002 (06.03.2002)	DE
102 21 930.3	16. Mai 2002 (16.05.2002)	DE
102 25 391.9	7. Juni 2002 (07.06.2002)	DE
102 27 795.8	21. Juni 2002 (21.06.2002)	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, TR, TT, TZ, UA, UG, US,
UZ, VN, YU, ZA, ZM, ZW.

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

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

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

(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

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 STRESS-ASSOCIATED GENETIC PRODUCTS FROM ASHBYA GOSSYPHII

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

(57) Abstract: The invention relates to novel polynucleotides from *Ashbya gossypii* oligonucleotides hybridised by the above; expression cassettes and vectors, containing said polynucleotides; micro-organisms transformed with the above; polypeptides coded by said polynucleotides and the use of the novel polypeptides and polynucleotides as targets to improve stress-resistance and, in particular, improve vitamin B2 production in micro-organisms of the species *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 Verbesserung der Stressresistenz und insbesondere der Verbesserung der Vitamin B2-Produktion in Mikroorganismen der Gattung *Ashbya*.



WO 03/014351 A2

BeschreibungNeue Stress-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 Verbesserung der Stressresistenz 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
- 25 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,
- 35

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

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

10

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

15 Mikroorganismen sind einer Reihe von Stressfaktoren ausgesetzt. Diese Faktoren lassen sich nach Ihrer Wirkung wie folgt nach Stresstypen einteilen:

- Nicht optimale Temperatur
- Oxidativer Stress
- Hyperosmolarität
- 20 • Hypoosmolarität
- Mechanischer Stress
- Nicht optimales Nährstoffangebot
- Kontakt mit Stoffen, die Giftcharakter aufweisen

25 Alle diese Stresstypen sind in der Lage, die Zelle zu schädigen oder sogar abzutöten. Um einen Schutz gegen diese Gefährdung aufzubauen, besitzen die Zellen Systeme der Stressantwort, die in der Lage sind, die Zelle an den Stress zu adaptieren, oder die Stressursache zu beseitigen.

30 Die Antwort der Zelle auf Stress beinhaltet Prozesse der Wachstumskontrolle, der Signaltransduktion, der Transkription und der posttranslationalen Kontrolle.

Die Nutzung von Mikroorganismen in Fermentern führt ebenfalls zu einer Reihe von Stresstypen (Attfield, P.V., Nature Biotechnology, 1997, 15, 1351 – 1357). So ist hier mit oxidativem Stress, einer nicht optimalen Osmolarität, mechanischen Stress (Scherstress hervorgerufen durch Rühren), einem nicht optimalen Nährstoffangebot und Kontakt mit giftigen Stoffwechselprodukten zu rechnen.

Aerob lebende Organismen haben eine reduzierte, zelluläre Redox-Umgebung aufrecht zu erhalten, parallel zu den prooxidativen Verhältnissen, die ein aerobes Leben mit sich bringt. Die unvollständige Reduktion von Sauerstoff zu Wasser führt zu der Bildung von Redox-aktiven Sauerstoffintermediaten (ROI), wie Superoxidradikalanionen, Hydrogenperoxid und Hydroxylradikale. ROI werden ebenfalls bei der β -Oxidation der Fettsäuren, bei Einfluß von ultravioletter oder radioaktiver Strahlung, durch Metalle und Redox-aktiven Substanzen gebildet. Oxidativer Stress resultiert aus abnormal hohen Mengen an ROIs, die den zellulären Redox-Status stören und zur Schädigung von Lipiden, Proteinen, DNA und RNA führen (Godon, C. et al. The Journal of Biological Chemistry, 1998, 273, 22480- 22489).

Aus diesem Grund prüfen Organismen ununterbrochen ihren Redoxstatus und adaptieren auf Änderungen durch die Induktion von Genen oder Aktivierung von Genprodukten, die in der Lage sind den zellulären Redox-Zustand wieder auf einen für die Zelle ungefährlichen Wert einzustellen. Gleichzeitig werden aber auch Gene oder Genprodukte aktiviert, die durch ROIs oder Strahlung geschädigte DNA erkennen und wieder reparieren können.

Gleiches gilt für die Adaption an veränderte Osmolarität. Auch hier prüft die Zelle den Osmostatus und reagiert auf Änderung durch Aktivierung von Genen und Genprodukten, die ungefährliche osmolarische Bedingungen einstellen (Blomberg, A, Electrophoresis, 1997, 18, 1429 – 1440). Entsprechendes gilt für mechanischen Stress und Stress durch Nährstofflimitierung.

Werden Organismen mit Giften konfrontiert, dann werden Gene oder Genprodukte aktiviert, die entweder das Gift durch Modifizierung unschädlich machen, oder Transportsysteme aktivieren, die in der Lage sind, die Gifte aus der Zelle zu schaffen (Baumeister, W, Critical Reviews in Microbiology, 1997, 23, 1 – 46).

Die Nutzung von Genen der Stressantwort zur Generierung von Mikroorganismen, vorzugsweise der Gattung *Ashbya*, insbesondere von *Ashbya gossypii* Stämmen, mit einer höheren Resistenz gegen Stress ist noch nicht beschrieben.

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

Gelöst wird obige Aufgabe durch Bereitstellung kodierenden Nukleinsäuresequenz, 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:

5

a) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Dioxygenase 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 70“ 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 „Oligo70 v“ trägt.

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

25 Die Inserts von „Oligo 70“ und „Oligo 70v“ besitzen mit dem MIPS Tag „YII057c“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 1 bzw. SEQ ID NO: 3. Eine vom korrespondierenden Gegenstrang zur SEQ ID NO: 1 bzw. vom kodierenden Strang der SEQ ID NO: 3 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Dioxygenase aus *S. cerevisiae*.

30

b) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion eines Hitzeschock-Proteins (insbesondere HSP26) 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 74“ trägt.

Ein Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 5, welche vorzugsweise aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar ist. Außerdem sind Gegenstand der Erfindung das dazu komplementäre Polynukleotid; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

Das Insert von „Oligo 74“ besitzt mit dem MIPS Tag „HSP26“ aus *S. cerevisiae* signifikante Homologien. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 5. Die vom korrespondierenden Gegenstrang abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem Hitzeschock-Protein (dem HSP26) aus *S. cerevisiae*.

c) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Methioninsulfoxid-Reduktase 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 113“ trägt.

Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 113v“ trägt.

Ein Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 8. Ein weiterer Gegenstand der Erfindung betrifft ein Polynukleotid, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 10 oder Fragmente davon. Vorzugsweise sind die Polynukleotide aus einem Mikroorganismus der Gattung *Ashbya*, insbesondere *A. gossypii* isolierbar. Außerdem sind Gegenstand der Erfindung die dazu komplementären Polynukleotide; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.

Die Inserts von „Oligo 113“ und „Oligo 113v“ besitzen mit dem MIPS Tag „Mxr1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 8 bzw. SEQ ID NO: 10. Die vom korrespondierenden Gegenstrang von SEQ ID NO: 8 bzw. von SEQ ID NO: 10 abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Methioninsulfoxid-Reduktase, einem für die Reduktion von Methioninsulfoxid verantwortlichen Protein aus *S. cerevisiae*.

- d) eine, vorzugsweise hochregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion eines DNA-Reparatur-Proteins kodiert.

5 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 60“ trägt.

10 Gemäß einer weiter bevorzugten Ausführungsform wurde erfindungsgemäß ein DNA-Klon isoliert, der für die Vollsequenz der erfindungsgemäßen Nukleinsäure kodiert und die interne Bezeichnung „Oligo 60v“ trägt.

15 Ein 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.

20 Die Inserts von „Oligo 60“ und „Oligo 60v“ besitzen mit dem MIPS Tag „Xrs2“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 12 bzw. SEQ ID NO: 14. Die vom kodierenden Strand abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einem DNA-Reparatur-Protein aus *S. cerevisiae*.

25

- e) eine, vorzugsweise niederregulierte, Nukleinsäuresequenz, welche für ein Protein mit der Funktion einer Komponente eines Chaperonin-haltigen T-Komplexes 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 161“ 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 161v“ trägt.

Ein Gegenstand der vorliegenden Erfindung betrifft ein Polynukleotid, umfassend eine Nuklein-

säuresequenz 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.

Die Inserts von „Oligo 161“ und „Oligo 161v“ besitzen mit dem MIPS Tag „Tcp1“ aus *S. cerevisiae* signifikante Homologien. Die Inserts besitzen eine Nukleinsäuresequenz gemäß SEQ ID NO: 16 bzw. SEQ ID NO: 18. Die vom korrespondierenden Gegenstrang (zu SEQ ID NO: 16) abgeleitete Aminosäuresequenz bzw. Aminosäureteilsequenz besitzt signifikante Sequenzhomologie mit einer Protein-Komponente eines Chaperonin-haltigen T-Komplexes aus *S. cerevisiae*.

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

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

Die Erfindung betrifft weiterhin Polypeptide bzw. Proteine, welche von den oben beschriebenen Polynukleotiden kodiert werden; sowie Peptidfragmente davon, welche eine Aminosäuresequenz aufweisen, die wenigstens 10 zusammenhängende Aminosäurereste gemäß SEQ ID NO: 2, 4, 6, 7, 9, 11, 13, 15, 17 oder 19 umfasst; sowie funktionale Äquivalente der erfindungsgemäßen Polypeptide bzw. Proteine.

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

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.

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

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 antigenbindende Fragmente 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
20 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 be-
25 schrieben.

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 Gegenstrang zu Position 115 bis 2 in SEQ ID NO:1) (obere Sequenz) und einer
35 Teilsequenz des MIPS-Tags YII057c aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 2 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 677 bis 399 in SEQ ID NO:5) (obere Sequenz) und einer Teilsequenz des MIPS-Tags HSP26 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 3 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem Gegenstrang zu Position 1259 bis 828 in SEQ ID NO:8) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Mxr1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Figur 4 zeigt ein Alignment zwischen einer erfindungsgemäßen Aminosäure-Teilsequenz (entsprechend dem kodierenden Strang zu Position 370 bis 1608 in SEQ ID NO:12) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Xrs2 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 558 bis 193 in SEQ ID NO:16) (obere Sequenz) und einer Teilsequenz des MIPS-Tags Tcp1 aus *S. cerevisiae* (untere Sequenz). Identische Sequenzpositionen sind zwischen den beiden Sequenzen angegeben. Ähnliche Sequenzpositionen sind mit „+“ gekennzeichnet.

Detaillierte Beschreibung der Erfindung:

Die erfindungsgemäßen Nukleinsäuremoleküle kodieren Proteine bzw. Proteine, die hier als Proteine der Stressantwort (z.B. mit Aktivität bezüglich DNA-Reparatur oder Entgiftung) oder kurz als „SA-Proteine“ bezeichnet werden. Diese SA-Proteine haben z.B. eine Funktion beim Schutz der Zelle vor schädlichen, äußeren Einflüssen (z.B. Radikalbildung, osmotischer Stress usw.) wie sie in Fermentern zu finden sind, insbesondere bei der Kultivierung von *A. gossypii*. Aufgrund der Verfügbarkeit von in *Ashbya gossypii* verwendbaren Klonierungsvektoren, wie z.B. offenbart in Wright und Philipsen (1991) Gene, 109, 99–105., und von Techniken zur genetischen Manipulation von *A. gossypii* und den verwandten Hefe-Arten lassen sich die erfindungsgemäßen Nukleinsäuremoleküle zur genetischen Manipulation dieser Organismen, insbesondere von *A. gossypii* verwenden, um sie als 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 auf einer indirekten Wirkung einer solchen Manipulation erfolgen.

Die vorliegende Erfindung beruht auf der Bereitstellung neuer Molekülen, die hier als SA-
5 Nukleinsäure und SA-Proteine bezeichnet werden und an der Stressantwort, insbesondere in *Ashbya gossypii*, (z.B. bei DNA-Reparatur oder Entgiftung) beteiligt sind. Die Aktivität der erfindungsgemäßen SA-Moleküle in *A. gossypii* beeinflusst die Vitamin B2-Produktion durch diesen Organismus. Vorzugsweise wird die Aktivität der erfindungsgemäßen SA-Moleküle so moduliert, dass die Stoffwechsel- und/oder Energiewege von *A. gossypii*, an denen die erfindungsgemä-
10 ßen SA-Proteine teilnehmen, hinsichtlich der Ausbeute, Produktion und/oder Effizienz der Vitamin B2-Produktion moduliert werden, was entweder direkt oder indirekt die Ausbeute, Produktion und/oder Effizienz der Vitamin B2-Produktion in *A. gossypii* moduliert.

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

Verbesserung der Vitamin B2-Produktion:

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

25 So kann durch eine effizientere Stressantwort, die Zelle robuster gegen äußere Einflüsse gemacht werden, so dass die Lebensfähigkeit und damit die Produktivität im Fermenter erhöht wird.

Die Mutagenese von einem oder mehreren erfindungsgemäßen SA-Proteinen kann auch zu SA-
30 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 SA-Proteine die Konzentration an freien Radikalen, welche Proteine angreifen und inaktivieren, verringern und so die Funktion von essentiellen Stoffwechselprozessen aufrecht erhalten. Zu diesen Prozessen gehört neben der Biosynthese des Produkts auch der Aufbau der
35 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) Biology of Prokaryotes, Thieme Verlag, Stuttgart). Durch Verbesserung von Wachstum und Vermehrung dieser veränderten Zellen ist es möglich, die Lebensfähig-

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

„Funktionale Äquivalente“ im obigen Sinne sind auch Präkursoren der beschriebenen Polypeptide sowie funktionale Derivate und Salze der Polypeptide. Unter dem Ausdruck „Salze“ versteht man sowohl Salze von Carboxylgruppen als auch Säureadditionssalze von Aminogruppen der erfindungsgemäßen Proteinmoleküle. Salze von Carboxylgruppen können in an sich bekannter Weise hergestellt werden und umfassen anorganische Salze, wie zum Beispiel Natrium-, Calcium-, Ammonium-, Eisen- und Zinksalze, sowie Salze mit organischen 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.

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

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

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

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

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

35 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.

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.

5

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. 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).

20

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

35

Im Stand der Technik sind mehrere Techniken zum Screening von Genprodukten kombinatorischer Banken, die durch Punktmutationen oder Verkürzung hergestellt worden sind, und zum Screening von cDNA-Banken auf Genprodukte mit einer ausgewählten Eigenschaft bekannt. Diese Techniken lassen sich an das schnelle Screening der Genbanken anpassen, die durch kombinatorische Mutagenese von erfindungsgemäßer Homologen erzeugt worden sind. Die am

häufigsten verwendeten Techniken zum Screening großer Genbanken, die einer Analyse mit hohem Durchsatz unterliegen, umfassen das Klonieren der Genbank in replizierbare Expressionsvektoren, Transformieren der geeigneten Zellen mit der resultierenden Vektorenbank und Exprimieren der kombinatorischen Gene unter Bedingungen, unter denen der Nachweis der gewünschten Aktivität die Isolation des Vektors, der das Gen codiert, dessen Produkt nachgewiesen wurde, erleichtert. Recursive-Ensemble-Mutagenese (REM), eine Technik, die die Häufigkeit funktioneller Mutanten in den Banken vergrößert, kann in Kombination mit den Screeningtests verwendet werden, um Homologe zu identifizieren (Arkin und Yourvan (1992) PNAS 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3):327-331).

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

Nukleinsäuresequenzen:

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

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 offenbarten vollständigen Sequenzen oder ein Abschnitt davon als Hybridisierungssonde 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 offenbarten Sequenzen oder einen Abschnitt davon, durch Polymerasekettenreaktion isolieren, wobei die Oligonukleotidprimer, die auf der Basis dieser Sequenz erstellt wurden, verwendet werden. Die so amplifizierte Nukleinsäure kann in einen geeigneten Vektor kloniert werden und durch DNA-Sequenzanalyse charakterisiert werden. Die erfindungsgemäßen Oligonukleotide, die einer 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, vorzugsweise mindestens etwa 25, wie z.B. etwa 40, 50 oder 75 aufeinanderfolgende Nukleotide eines Sense-Stranges einer erfindungsgemäßen Nukleinsäuresequenz oder eines entsprechenden Antisense-Stranges hybridisiert.

Weitere erfindungsgemäße Nukleinsäuresequenzen sind abgeleitet von den Sequenzen gemäß SEQ ID NO: 1, 3, 5, 8, 10, 12, 14, 16 und 18 und unterscheiden sich davon durch Addition, Substitution, Insertion oder Deletion einzelner oder mehrerer Nukleotide, kodieren aber weiterhin für Polypeptide mit dem gewünschten Eigenschaftsprofil.

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

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

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

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

Ein weiterer Aspekt der Erfindung betrifft „Antisense“-Nukleinsäuren. Diese umfaßt eine Nukleotidsequenz, die zu einer kodierenden „Sense“-Nukleinsäure, komplementär ist. Die Antisense-Nukleinsäure kann zum gesamten kodierenden Strang oder nur zu einem Abschnitt davon komplementär sein. Bei einer weiteren Ausführungsform ist das Antisense-Nukleinsäuremolekül antisense zu einem nicht-kodierenden Bereich des kodierenden Stranges einer Nukleotidsequenz.

Der Begriff "nicht-kodierender Bereich" betrifft die als 5'- und 3'-untranslatierte Bereiche bezeichneten Sequenzabschnitte.

Ein Antisense-Oligonukleotid kann bspw. etwa 5, 10, 15, 20, 25, 30, 35, 40, 45 oder 50 Nukleotide lang sein. Eine erfindungsgemäße Antisense-Nukleinsäure kann durch chemische Synthese und enzymatische Ligationsreaktionen mittels im Fachgebiet bekannter Verfahren konstruiert werden. Eine Antisense-Nukleinsäure kann chemisch synthetisiert werden, wobei natürlich vorkommende Nukleotide oder verschieden modifizierte Nukleotide verwendet werden, die so gestaltet sind, daß sie die biologische Stabilität der Moleküle erhöhen, oder die physikalische Stabilität des Duplexes erhöhen, der zwischen der Antisense- und Sense-Nukleinsäure entstanden ist. Beispielsweise können Phosphorthioat-Derivate und acridinsubstituierte Nukleotide verwendet werden. Beispiele modifizierter Nukleotide, die zur Erzeugung der Antisense-Nukleinsäure verwendet werden können, sind u.a. 5-Fluoruracil, 5-Bromuracil, 5-Chloruracil, 5-Ioduracil, Hypoxanthin, Xanthin, 4-Acetylcytosin, 5-(Carboxyhydroxymethyl)uracil, 5-Carboxymethylaminomethyl-2-thiouridin, 5-Carboxymethylaminomethyluracil, Dihydrouracil, Beta-D-Galactosylqueosin, Inosin, N6-Isopentenyladenin, 1-Methylguanin, 1-Methylinosin, 2,2-Dimethylguanin, 2-Methyladenin, 2-Methylguanin, 3-Methylcytosin, 5-Methylcytosin, N6-Adenin, 7-Methylguanin, 5-Methylaminomethyluracil, 5-Methoxyaminomethyl-2-thiouracil, Beta-D-Mannosylqueosin, 5'-Methoxycarboxymethyluracil, 5-Methoxyuracil, 2-Methylthio-N6-isopentenyladenin, Uracil-5-oxyessigsäure (v), Wybutoxosin, Pseudouracil, Queosin, 2-Thiocytosin, 5-Methyl-2-thiouracil, 2-Thiouracil, 4-Thiouracil, 5-Methyluracil, Uracil-5-oxyessigsäuremethylester, Uracil-5-oxyessigsäure (v), 5-Methyl-2-thiouracil, 3-(3-Amino-3-N-2-carboxypropyl)uracil, (acp3)w und 2,6-Diaminopurin. Die Antisense-Nukleinsäure kann auch biologisch hergestellt werden, indem ein Expressionsvektor verwendet wird, in den eine Nukleinsäure in Antisense-Richtung subkloniert worden ist.

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

Das Antisense-Molekül kann so modifiziert werden, daß es spezifisch an einen Rezeptor oder an ein Antigen bindet, das auf einer ausgewählten Zelloberfläche exprimiert wird, z.B. durch Verknüpfen des Antisense-Nukleinsäuremoleküls mit einem Peptid oder einem Antikörper, das/der an einen Zelloberflächenrezeptor oder Antigen bindet. Das Antisense-Nukleinsäuremolekül kann auch unter Verwendung der hier beschriebenen Vektoren an Zellen verabreicht werden. Zur Erzielung hinreichender intrazellulärer Konzentrationen der Antisense-Moleküle sind Vektorkon-

strukture, in denen sich das Antisense-Nukleinsäuremolekül unter der Kontrolle eines starken bakteriellen, viralen oder eukaryotischen Promotors befindet, bevorzugt.

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

Gegenstand der Erfindung sind auch Ribozyme. Dies sind katalytische RNA-Moleküle mit Ribonukleaseaktivität, die eine einzelsträngige Nukleinsäure, wie eine mRNA, zu der sie einen komplementären Bereich haben, spalten können. Somit können Ribozyme (z.B. Hammerhead-Ribozyme (beschrieben in Haselhoff und Gerlach (1988) Nature 334:585-591)) zur katalytischen Spaltung von erfindungsgemäßen Transkripten verwendet werden, um dadurch die Translation der entsprechenden Nukleinsäure zu hemmen. Ein Ribozym mit Spezifität für eine erfindungsgemäße kodierende Nukleinsäure kann z.B. auf der Basis einer hierin konkret 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 z.B. Bartel, D., und Szostak, J.W. (1993) Science 261:1411-1418).

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

Expressionskonstrukte und Vektoren:

Gegenstand der Erfindung sind außerdem Expressionskonstrukte, enthaltend unter der genetischen Kontrolle regulativer Nukleinsäuresequenzen eine für ein erfindungsgemäßes Polypeptid kodierende Nukleinsäuresequenz; sowie Vektoren, umfassend wenigstens eines dieser Expres-

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

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

Beispiele für brauchbare Promotoren sind: cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, lacIq-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, lambda-PR- oder im lambda-PL-Promotor, die vorteilhafterweise in gram-negativen Bakterien Anwendung finden; sowie die gram-positiven Promotoren amy und SPO2, die Hefepromotoren ADC1, MFalpha, 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 P₁P₁-Promotor. Prinzipiell können alle natürlichen Promotoren mit ihren Regulationssequenzen verwendet werden. Darüber hinaus können auch synthetische Promotoren vorteilhaft verwendet werden.

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.

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
5 auch eine Verstärkung der Translation möglich, indem beispielsweise die Stabilität der mRNA verbessert wird.

Die Herstellung einer Expressionskassette erfolgt durch Fusion eines geeigneten Promotors mit einer geeigneten Monooxygenase-Nukleotidsequenz sowie einem Terminator- oder Polyadenylierungssignal. Dazu verwendet man gängige Rekombinations- und Klonierungstechniken, wie
10 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,
15 gy, Greene Publishing Assoc. and Wiley Interscience (1987) beschrieben sind.

Das rekombinante Nukleinsäurekonstrukt bzw. Genkonstrukt wird zur Expression in einem geeigneten Wirtsorganismus vorteilhafterweise in einen wirtsspezifischen Vektor inseriert, der eine 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
20 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.
25

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.
30

Nicht-Fusionsprotein-Expressionsvektoren wie pTrc (Amann et al., (1988) Gene 69:301-315) und pET 11d (Studier et al. Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Kalifornien (1990) 60-89).
35

- Hefe-Expressionsvektor zur Expression in der Hefe *S. cerevisiae*, wie pYepSec1 (Baldari et al., (1987) Embo J. 6:229-234), pMFa (Kurjan und Herskowitz (1982) Cell 30:933-943), pJRY88 (Schultz et al. (1987) Gene 54:113-123) sowie pYES2 (Invitrogen Corporation, San Diego, CA). Vektoren und Verfahren zur Konstruktion von Vektoren, die sich zur Verwendung in anderen
- 5 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.
- 10 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.,
- 15 Kemper, E., Schell, J. und Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", Plant Mol. Biol. 20:1195-1197; und Bevan, M.W. (1984) "Binary Agrobacterium vectors for plant transformation", Nucl. Acids Res. 12:8711-8721.
- Säugetier-Expressionsvektoren, wie pCDM8 (Seed, B. (1987) Nature 329:840) und pMT2PC
- 20 (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,
- 25 Cold Spring Harbor, NY, 1989 beschrieben.

Rekombinante Mikroorganismen:

- Mit Hilfe der erfindungsgemäßen Vektoren sind rekombinante Mikroorganismen herstellbar, welche beispielsweise mit wenigstens einem erfindungsgemäßen Vektor transformiert sind und zur
- 30 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,
- 35 Protoplastenfusion, Elektroporation, retrovirale Transfektion und dergleichen, verwendet, um die genannten Nukleinsäuren im jeweiligen Expressionssystem zur Expression zu bringen. Geeignete Systeme werden beispielsweise in Current Protocols in Molecular Biology, F. Ausubel et al., Hrsg., Wiley Interscience, New York 1997, oder Sambrook et al. Molecular Cloning: A Laboratory

Manual. 2. Aufl., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989 beschrieben.

Erfindungsgemäß sind auch homolog rekombinierten Mikroorganismen herstellbar. Dazu wird
5 ein Vektor hergestellt, der zumindest einen Abschnitt eines erfindungsgemäßen Gens oder einer
kodierenden Sequenz enthält, worin gegebenenfalls wenigstens eine Aminosäure-Deletion, -
Addition oder -Substitution eingebracht worden ist, um die erfindungsgemäße Sequenz zu ver-
ändern, z.B. funktionell zu disruptieren ("Knockout"-Vektor). Die eingebrachte Sequenz kann
z.B. auch ein Homologes aus einem Verwandten Mikroorganismus sein oder aus einer Säuge-
10 tier-, Hefe- oder Insektenquelle abgeleitet sein. Der zur homologen Rekombination verwendete
Vektor kann alternativ derart ausgestaltet sein, daß das endogene Gen bei homologer Rekombi-
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 SA-Gens ist
15 im homologen Rekombinationsvektor. Die Konstruktion geeigneter Vektoren zur homologen Re-
kombination 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 erfin-
dungsgemäßen Nukleinsäuren, ihrer Allelvarianten, ihrer funktionellen Äquivalente oder Derivate
20 ermöglichen. Unter Wirtsorganismen sind beispielsweise Bakterien, Pilze, Hefen, pflanzliche
oder tierische Zellen zu verstehen. Bevorzugte Organismen sind Bakterien, wie solche der Gat-
tungen *Escherichia*, wie z. B. *Escherichia coli*, *Streptomyces*, *Bacillus* oder *Pseudomonas*, euka-
ryotische Mikroorganismen, wie *Saccharomyces cerevisiae*, *Aspergillus*, höhere eukaryotische
Zellen aus Tieren oder Pflanzen, beispielsweise Sf9 oder CHO-Zellen. Bevorzugte Organismen
25 sind aus der Gattung *Ashbya*, insbesondere aus *A. gossypii*-Stämmen ausgewählt.

Die Selektion erfolgreich transformierter Organismen kann durch Markergene erfolgen, die eben-
falls im Vektor oder in der Expressionskassette enthalten sind. Beispiele für solche Markergene
sind Gene für Antibiotikaresistenz und für Enzyme, die eine farbgebende Reaktion katalysieren,
30 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. Markerprotei-
ne, die an der Zelloberfläche präsentiert werden, können zur Selektion mittels Affinitätschroma-
35 tographie 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 Säugertierzellen, wie CHO-Zellen, und Vektoren, wie pcDNA3neo-Vektor, die für Säugertierzellen 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.

10 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.

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

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

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

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

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

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

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

Folgende nichtlimitierende Beispiele beschreiben spezielle Ausführungsformen der Erfindung.

Allgemeine experimentelle Angaben

a) Allgemeine Klonierungsverfahren

Die im Rahmen der vorliegenden Erfindung durchgeführten Klonierungsschritte wie z.B. Restriktionsspaltungen, Agarose Gelelektrophorese, Reinigung von DNA-Fragmenten, Transfer von Nukleinsäuren auf Nitrozellulose und Nylonmembranen, Verknüpfen von DNA-Fragmenten,

Transformation von E. coli Zellen, Anzucht von Bakterien, Vermehrung von Phagen und Sequenzanalyse rekombinanter DNA wurden wie bei Sambrook et al. (1989) a.a.O. beschrieben durchgeführt.

5 b) Polymerasekettenreaktion (PCR)

PCR wurde nach Standardprotokoll mit folgendem Standardansatz durchgeführt:

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

c) Kultivierung von E.coli

15 Die Kultivierung von rekombinanten E. coli-Stämme DH5α wurde in LB-Amp Medium (Trypton 10,0g, NaCl 5,0 g, Hefeextrakt 5,0 g, Ampicillin 100 g/ml H₂O ad 1000 ml) bei 37 °C kultiviert. Dazu wurde jeweils eine Kolonie mittels Impföse von einer Agarplatte in 5 ml LB-Amp überführt. Nach ca. 18 h Stunden Kultivierung bei einer Schüttelfrequenz von 220 Upm wurden 400 ml Medium in einem 2-l-Kolben mit 4 ml Kultur inokuliert. Die Induktion der P450-Expression in E. coli
20 erfolgte nach Erreichen eines OD₅₇₈-Wertes zwischen 0,8 und 1,0 durch eine drei- bis vierstündige Hitzeschockinduktion bei 42 °C.

d) Reinigung des gewünschten Produktes aus der Kultur

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

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

35 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:

5 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.

10

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.

15

20

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.

25

30

Auf diese Weise wurde ein Klon identifiziert, der ein Insert mit der internen Bezeichnung „Oligo 70“ trägt und mit dem MIPS Tag „YII057c“ aus *S. cerevisiae* signifikante Homologien besitzt. Das Insert besitzt eine Nukleinsäuresequenz gemäß SEQ ID NO: 1.

35

Beispiel 5:

Auswertung der Sequenzdaten mit Hilfe einer BLASTX Suche

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

Die entsprechenden Alignments sind in beiliegenden Figuren 1 bis 5 gezeigt.

a) Die vom korrespondierenden Gegenstrang zu SEQ ID NO: 1 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Dioxygenase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 115 bis 2 aus SEQ ID NO:1) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 1 dargestellt. SEQ ID NO: 2 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz.

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit die Funktion einer Dioxygenase zugeordnet werden.

20

b) Die vom korrespondierenden Gegenstrang zu SEQ ID NO: 5 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem Hitzeschock-Protein aus *S. cerevisiae*. Eine davon abgeleitete Aminosäure-Teilsequenz (entsprechend den Nucleotiden 677 bis 399 aus SEQ ID NO:5) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 2 dargestellt. SEQ ID NO: 6 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz. SEQ ID NO: 7 zeigt eine weitere Teilsequenz.

25

Der ermittelten *A. gossypii* Nukleinsäuresequenz konnte damit die Funktion eines Hitzeschock-Proteins zugeordnet werden.

30

c) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:8 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einer Methioninsulfoxid-Reduktase aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 1259 bis 828 aus SEQ ID NO:8) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 3 dargestellt. SEQ ID NO: 9 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

35

Der ermittelten *A. gossypii* Nukleinsäuresequenz konnte damit die Funktion einer Methioninsulfoxid-Reduktase zugeordnet werden.

- d) Die vom kodierenden Strang zu SEQ ID NO:12 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit einem DNA-Reparatur-Protein aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 370 bis 1608 aus SEQ ID NO:12) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 4 dargestellt. SEQ ID NO: 13 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

Der ermittelten *A. gossypii* Nukleinsäuresequenz konnte damit die Funktion eines DNA-Reparatur-Proteins zugeordnet werden.

10

- e) Die vom korrespondierenden Gegenstrang zu SEQ ID NO:16 abgeleitete Aminosäuresequenz besitzt signifikante Sequenzhomologie mit eines Chaperonin-haltigen T-Komplexes aus *S. cerevisiae*. Eine davon abgeleiteten Aminosäure-Teilsequenz (entsprechend den Nucleotiden 558 bis 193 aus SEQ ID NO:16) mit einer Teilsequenz des *S. cerevisiae* Enzyms ist in Figur 5 dargestellt. SEQ ID NO: 17 zeigt eine N-terminal verlängerte Aminosäure-Teilsequenz .

15

Die ermittelte *A. gossypii* Nukleinsäuresequenz konnte damit die Funktion eines Chaperonin-haltigen T-Komplexes zugeordnet werden.

20 **Beispiel 6:**

Isolierung der Full-Length-DNA

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

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

30

- 35 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 *XbaI* und Dephosphorylierung mit alkalischer Kalbsdarm-Phosphatase (Boeh-

ringer) geöffnet, gefolgt von einem Öffnen der Klonierungsstelle mit *Bam*HI. Die Ligationen wurden über Nacht bei 15 °C in 20 µl, enthaltend 2,5 µg teilverdauter chromosomaler DNA, 1 µg Super-Cos1-Vektorenarme, 40 mM Tris-HCl, pH 7,5, 10 mM MgCl₂, 1 mM Dithiothreitol, 0,5 mM ATP und 2 Weiss-Einheiten T4-DNA-Ligase (Boehringer) durchgeführt. Die Ligationsprodukte wurden unter Verwendung der Extrakte und des Protokolls von Stratagene (Gigapack II Packaging Extract) in vitro verpackt. Das verpackte Material wurde zur Infizierung von *E. coli* NM554 (*recA13*, *araD139*, Δ (*ara*,*leu*)7696, Δ (*lac*)17A, *galU*, *galK*, *hsrR*, *rps*(*str*^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

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 dupliziert. Vor dem Einfrieren und nach dem Auftauen (vor irgendwelchen anderen Maßnahmen) wurden die Platten kurz in einem Mikrotiterschüttler (Infors) geschüttelt, um eine homogene Zellsuspension zu gewährleisten. Mittels eines Robotersystems (Bio-Robotics), mit dem geringe Mengen an Flüssigkeit aus 96 Vertiefungen einer Mikrotiterplatte auf Nylonmembran (GeneScreen Plus, New England Nuclear) transferiert werden können, wurden einzelne Klone auf Nylonmembranen platziert. Nach dem Transfer der Kultur aus den 96-er Mikrotiterplatten (1920 Klone) wurden die Membranen auf die Oberfläche von LB-Agar mit Ampicillin (50 µg/ml) in 22 x 22 cm Kulturschalen (Nunc) platziert und über Nacht bei 37 °C inkubiert. Vor Erreichen der Zellkonfluenz wurden die Membranen, wie von Herrmann, B. G., Barlow, D. P. und Lehrach, H. (1987) in Cell 48, S. 813-825 beschrieben, prozessiert, wobei als zusätzliche Behandlung nach dem ersten Denaturierungsschritt ein 5-minütiges Bedampfen der Filter auf einem in Denaturierungslösung getränkten Pad über einem kochenden Wasserbad hinzukommt.

Mit Hilfe des Random-Hexamer-Primer-Verfahrens (Feinberg, A. P. und Vogelstein, B. (1983), Anal. Biochem. 132, S. 6-13) wurden doppelsträngige Sonden durch Aufnahme von [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 10⁶ cpm/ml) hybridisiert. Typischerweise wurden

Waschschritte 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
5 bei 95 °C für 2 x 20 min in 2 mM Tris-HCl, pH 8,0, 0,2 mM EDTA, 0,1 % SDS gestrippt.

c) Zurückgewinnung positiver Kolonien aus der aufbewahrten Genbank

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

15 d) Full-Length DNA

Auf die oben beschriebenen Weise konnten die im folgendenden genannten Klone identifiziert werden, der ein Insert mit der Vollsequenz zu den gemäß Beispiel 5 erhaltenen Teilsequenzen tragen. Diese Klon tragen die internen Bezeichnungen „Oligo 70v“, „Oligo 113v“, „Oligo 60v“ und
20 „Oligo 161v“. Eine Übersicht über sämtliche erfindungsgemäße Teil- und Vollsequenzen ist in folgender Tabelle 1 zu finden:

Tabelle 1: Sequenzübersicht

<u>SEQ ID NO:</u>	<u>Oligo</u>	<u>Beschreibung der Sequenz</u>	<u>Sequenzhomologie</u>
1	070	DNA-Teilsequenz	Dioxygenase aus <i>S. cerevisiae</i>
2	070	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 1	
3	070	DNA-Full-lenght-Sequenz	
4	070	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 3 von Position 402 bis 1511	
5	074	DNA-Teilsequenz	Hitzeschock-Proteins aus <i>S. cerevisiae</i>
6	074	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 5	
7	074	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 5	
8	113	DNA-Teilsequenz	Methioninsulfoxid-Reduktase aus <i>S. cerevisiae</i>
9	113	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 8	
10	113	DNA-Full-lenght-Sequenz	
11	113	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 10 von Position 519 bis 1310	
12	060	DNA-Teilsequenz	DNA-Reparatur-Protein aus <i>S. cerevisiae</i> .
13	060	Aminosäureteilsequenz, abgeleitet vom kodierenden Strang der SEQ ID NO: 12	
14	060	DNA-Full-lenght-Sequenz	
15	060	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 14 von Position 570 bis 3176	
16	161	DNA-Teilsequenz	Protein-Komponente eines Chaperonin-haltigen T-Komplexes aus <i>S. cerevisiae</i> .
17	161	Aminosäureteilsequenz, abgeleitet vom Gegenstrang zu SEQ ID NO: 16	
18	161	DNA-Full-lenght-Sequenz	
19	161	Aminosäuresequenz, entsprechend dem kodierenden Bereich der SEQ ID NO: 18 von Position 479 bis 2152	

Patentansprüche

1. Polynukleotid, welches für ein Protein der Stressantwort kodiert und aus *Ashbya gossypii* isolierbar ist.
2. Polynukleotid nach Anspruch 1, welches für ein Protein der Stressantwort kodiert, das ausgewählt ist unter Proteinen mit der biologischen Aktivität einer Dioxygenase, eines Hitzeschockproteins, einer Methioninsulfoxid Reduktase, eines DNA-Repair-Proteins und eines Proteins des Chaperonin-haltigen T-Komplexes.
3. Polynukleotid, nach Anspruch 1 oder 2, umfassend eine Nukleinsäuresequenz gemäß SEQ ID NO: 1, 5, 8, 12 oder 16, welche vorzugsweise aus *Ashbya gossypii* isolierbar ist, oder Fragmente davon; das dazu komplementäre Polynukleotid; und die von diesen Polynukleotiden durch Entartung des genetischen Codes abgeleiteten Sequenzen.
4. Polynukleotid nach Anspruch 3, welches eine Nukleinsäuresequenz gemäß SEQ ID NO: 3, 10, 14 oder 18 oder ein Fragment davon umfasst.
5. Oligonukleotid, welches mit einem Polynukleotid nach einem der vorhergehenden Ansprüche, insbesondere unter stringenten Bedingungen, hybridisiert.
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.
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 Aminosäurereste gemäß SEQ ID NO: 2, 4, 6, 7, 9, 11, 13, 15, 17 oder 19 umfasst; sowie funktionale Äquivalente davon, insbesondere solche welche eine der in Anspruch 2 definierten Aktivitäten besitzen.

8. Expressionskassette, umfassend in operativer Verknüpfung mit wenigstens einer regulativen Nukleinsäuresequenz eine Nukleinsäuresequenz nach einem der Ansprüche 1 bis 6.
- 5 9. Rekombinanter Vektor, umfassend wenigstens eine Expressionskassette nach Anspruch 8.
10. Prokaryotischer oder eukaryotischer Wirt, transformiert mit wenigstens einem Vektor nach Anspruch 9.
- 10 11. Prokaryotischer oder eukaryotischer Wirt, in welchem die funktionale Expression wenigstens eines Gens moduliert ist, das für ein Polypeptid nach Anspruch 7 kodiert; oder in welchem die biologische Aktivität eines Polypeptids nach Anspruch 7 erniedrigt oder erhöht ist.
- 15 12. Wirt nach Anspruch 10 oder 11 aus der Gattung *Ashbya*.
13. Verwendung einer Expressionskassette nach Anspruch 8, eines Vektors nach Anspruch 7 oder eines Wirts nach einem der Ansprüche 10 bis 12 zur mikro-
20 biologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon.
14. Verwendung einer Expressionskassette nach Anspruch 8, eines Vektors nach Anspruch 9 oder eines Wirts nach einem der Ansprüche 10 bis 12 zur rekombinanten
25 Herstellung eines Polypeptids nach Anspruch 7.
15. Verfahren zum Nachweis eines Effektortargets für die Modulation der mikro-
biologischen Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon, wobei man einen Mikroorganismus, der zur mikrobiologischen Produktion von
30 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, insbesondere bindet, den Einfluss des Effektors auf die Menge des mikrobiologisch produzierten Vitamins B2, und/oder des Präkursors und/oder eines Derivats davon
35 validiert; und das Target gegebenenfalls isoliert.

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.
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 7 wechselwirken;
 - c) niedermolekularen Effektoren, welche die biologische Aktivität eines Polypeptids nach Anspruch 7 modulieren;
 - d) Antisense-Nukleinsäuresequenzen.
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.
19. Verfahren nach Anspruch 18, wobei man vor und/oder während der Kultivierung des Wirts diesen mit einem Effektor nach Anspruch 17 behandelt.
20. Verfahren nach Anspruch 18 oder 19, wobei der Wirt ausgewählt ist unter Mikroorganismen der Gattung *Ashbya*.
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 bis 4 und 6 oder eines Polypeptids nach Anspruch 7 als Target zur Modulation der Produktion von Vitamin B2 und/oder Präkursoren und/oder Derivaten davon in einem Mikroorganismus der Gattung *Ashbya*.

23. Verwendung eines Polynukleotids nach einem der Ansprüche 1 bis 4 und 6 oder
eines Polypeptids nach Anspruch 7 als Target zur Modulation der Stressantwort oder
eines stressbedingten Folgezustandes in einem Mikroorganismus der Gattung
5 *Ashbya* während der Kultivierung zur mikrobiologischen Produktion von Vitamin B2
und/oder Präkursoren und/oder Derivaten davon.

24. Wirt nach Anspruch 12 mit verbesserter Resistenz gegen insbesondere
fermentativen Stress.

Fig. 1**Oligo 70**

```
115 FDTHFFAEQDEVGDDGILRLANRYILNSKFPKFLPTWD 2   A. gossypii
    FDTHF+  QDEV  G+L++  Y   SK+P +LPTWD
40  FDTHFYDQGDEVSPSGLLKIRKSYREKSKYPDYLPTWD 77  S. cerevisiae
```

Fig. 2**Oligo74**

```
677 GHRXQDNINIEFHSDXXELVISGEVXXXXXXXXXXKNGWRVRERATGSFRRVIGLPDKPGVD 498
      G + + +I+IE+H + +++++SGE+          K+ +V+E ++G F+RVI LPD PGVD
112 GVKSKKDIDIEYHQNKQILVSGEIPSTLNEESKDKVKVKESSGKFKRVITLPDYPGVD 171

497 AEHITANYEHGVLTLRVPKLELRAD--QAVRKIQI 399      A. gossypii
      A++I A+Y +GVLTL VPKL+ + D      V+KI++
172 ADNIAKADYANGVLTLTVPKLKPQKDGKNHVKKIEV 206      S. cerevisiae
```

Fig. 3**Oligo 113**

1259 IVDLVAGYANGNVPTDTKGSHIX-----SEFV*VTPISQRWSRSPFDPEVLGLSELC 1104
IVD GYANG S + ++F V +S ++P+V+ L EL
41 IVDCKVGYANGEESKKDSPSSVSYKRVCGGDTDFAEVLQVS-----YNPKVITLRELT 93

1103 DFFFRMHDPTTLNAQGPDVGTQYQQCSCMRIPDEDLSQVRQILERWQPRWANCITEVAL 924
DFFFR+HDPTT N+QGPD GTQY+ D DL ++ +I E WQP+W N I T +
94 DFFFRIHDP TT SNSQGPDKGTQYRS -GLFAHSDADLKELAKIKEEWQPKWGNKIATVIEP 152

923 AESFYDAEEYHQRYLTNNPTGYQCPTHYLRNV 828 A. gossypii
++FYDAEEYHQ YL NP GY CPTHYLR +
153 IKNFYDAEEYHQLYLDKNPQGYACPTHYLR 184 S. cerevisiae

Fig. 4

Oligo 60

370 MWVFKCNYQKSGQLGALQNISALCPEREFIVGRASKSTLVIE-EKGVSRTHL--RLEVI 540
 MWV + YQ + + G++ ISC L + + +GR+SK+ L+I+ +K +SR H+ + E+
 1 MWVVR--YQNTLEDGSISFISCCLOAFKTYSIGRSSKNPLIIKNDKSISRQHITFKWEIN 58

 541 DD----HVNILL--GSQLKVGGEDASKGSVLPGPSQ--SPVILEVGSHSVQCILEWAQ 696
 + H ++ L+ G + + G S I+E+G+ ++ EW
 59 NSSDLKHSSLCLVNKGKLTSLNKKFMKVGETFTINASDVLKSTIIELGTTPIRIEFWIN 118

 697 WEFKVPNTLLQEDPALLEVIGLIRVVTSFSKATTHHIIRPHEDDQDYE-KYLFALVKG 873
 + +P L Q L E GI S + + +I + +D + L+ALV
 119 EVWNIPPHLTQFRTMLSE----YGISTEISINDIPANLMISDYPKSEDNSIRELYALVST 174

 874 IPI-----LHLEFLKQFTALLRQP---VTDFDPQLRSIQEQHYAFRG 990
 IP+ +L+F + + ++ P V DFDP +
 175 IPMKKSRFLMELCNTLLPTSKTNLKFDEMWNMISNPEYNVDFDPNI----- 222

 991 FTKKAADLNGLHFIVNRKHDFEMLKYTLEIGGGTVHFCHDINDLSACLKTIPINGALALK 1170
 K LN + + K + LS+ L+T IN
 223 LLSKFMRLNNIRVLTTIKSE-----PRLSSLLRTFNIN---LFA 258

 1171 YDHGISA EVKKVD---DTRYCEVL---HSHGIQLFTVHELT KAILYDKLAQLLSKRPP 1326
 +D+ I + K VD T Y + +G L T+ + +I+ L+ +++ +
 259 FDN-IDS LYKYVDSLEASTEYLILTTTDDKKENGKILCTIKTMLTSIIDGTL SAVINMKGA 317

 1327 TMMKL-----VEPVPSTISQTQEMKQESLSPISLGNTTIKDLPRKDKTRKRLKKPQPL 1485
 + L + STI +T + SP+ + +K K +R + PL
 318 SSRTLDNGKFDQISEGMSTILKTSRAPEVEASPV-----VSKKRKLNRR--RVLPL 366

 1486 DSLSFFAGGLSRQNSSIGGTADLVKQESMSLGADTHYGSPN 1608 *A. gossypii*
 DSL FFAGGLS + S + K+ + + T SPN
 367 DSLDFFAGGLSTKTLSENRSLTDAKRLNCGAESKTVISSPN 407115 *S. cerevisiae*

Fig. 5**Oligo161**

558 TTVGSREQLAIAEFAAAMLVIPKTLAVNAAKDSSDLVAKLRSYHAASQQAKPDDLKKRKY 379
TTVGSREQLAIAEFAAA+L+IPKTLAVNAAKDSS+LVAKLRSYHAASQ AKP+D+K+R Y
438 TTVGSREQLAIAEFAAALLIIPKTLAVNAAKDSSSELVAKLRSYHAASQMAKPEDVKRRSY 497

378 RNYGLDLIRGKVDEVGAGVLEPTIXXXXXXXXXSALEACIAILRIDTMITVDPEPPKEDPH 199
RNYGLDLIRGK+VDE+ AGVLEPTI SALEAC+AILRIDTMITVDPEPPKEDPH
498 RNYGLDLIRGKIVDEIHAGVLEPTISKVKSLKSALEACVAILRIDTMITVDPEPPKEDPH 557

198 AH 193 A. gossypii
H
558 DH 559 S. cerevisiae

SEQUENCE LISTING

<110> BASF Aktiengesellschaft

<120> Neue Stress-assoziierte Genprodukte aus Ashbya Gossypii

<130> M/42237/PCT

<160> 19

<170> PatentIn version 3.1

<210> 1

<211> 522

<212> DNA

<213> Ashbya gossypii

<220>

<221> gene

<222> (2) .. (115)

<223>

<220>

<221> misc_feature

<223> Oligo 70

<400> 1

gatcccaagt tggaagaaat ttgggggaact tcgaatttaa aatataacga tttgccaatc	60
ttaatatattcc gtcgtcgcca acctcatcct gttcagcaaa aaaatgggta tcaaaattac	120
cattagcacc attcccagct cctactctaa ctgtcatctt gtctagtcta ctctttaaaa	180
tgaaggccgg caactactgg gcttacatca gacttatata tctgcacgta ttacatat	240

```

gatctgtggt tggtttgtct tgcgatagct agtaactcat aacaaagtat tccgaccatg      300
gctctctgtg tcttccatgc agacacggag actcgggtgct tcttattgct cctccgcatg      360
aacgtgacag cggtcaccgt cttcaagata gcagtattat cacgtgacct atatatccac      420
aattacctac aggaaattat tctgtgggct aagtcttcag taatagacct gctaagcaac      480
tacgcagcag tgctgggtgtt gcgtgcgttg aacgcaggag cc                          522

```

<210> 2

<211> 40

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 70

<400> 2

```

Gly Asn Phe Asp Thr His Phe Phe Ala Glu Gln Asp Glu Val Gly Asp
1           5           10           15

```

```

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

```

```

Pro Lys Phe Leu Pro Thr Trp Asp
          35           40

```

<210> 3

<211> 2127

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (402)..(1511)

<223>

<223> Oligo 70

<400> 3

gggtcctgcg	ttcaacgcac	gcaacaccag	cactgctgcg	tagttgctta	gcaggcttat		60
tactgaagac	ttagcccaca	gaataatttc	ctgtaggtaa	ttgtggatat	ataggtcacg		120
tgataatact	gctatcttga	agacggtgac	cgctgtcacg	ttcatgcgga	ggagcaataa		180
gaagcaccga	gtctccgtgt	ctgcatggaa	gacacagaga	gccatggctg	gaatactttg		240
ttatgagtta	ctagctatcg	caagacaaaac	caaccacaga	tcaaatatgt	aatacgtgca		300
gatatataag	tctgatgtag	gcccagtagt	tgccggcctt	cattttaaag	agtagactag		360
acaagatgac	agttagagta	ggagctggga	atgggtgctaa	t ggt aat ttt gat acc Gly Asn Phe Asp Thr 1 5			416
cat ttt ttt gct gaa cag gat gag gtt ggc gac gac gga ata tta aga His Phe Phe Ala Glu Gln Asp Glu Val Gly Asp Asp Gly Ile Leu Arg 10 15 20							464
ttg gca aat cgt tat att tta aat tcg aag ttc ccc aaa ttt ctt cca Leu Ala Asn Arg Tyr Ile Leu Asn Ser Lys Phe Pro Lys Phe Leu Pro 25 30 35							512
act tgg gat cat tcg caa aag tac ccg ccc tta gta ttc cat aag tat Thr Trp Asp His Ser Gln Lys Tyr Pro Pro Leu Val Phe His Lys Tyr 40 45 50							560
ttg gat cct gct ctg aga gca gat cct cag tta ccg agg ttt ttg aat Leu Asp Pro Ala Leu Arg Ala Asp Pro Gln Leu Pro Arg Phe Leu Asn 55 60 65							608
gag aac gtc caa att tct aaa ata acg ccc cgg ttc gga agc gag gtt Glu Asn Val Gln Ile Ser Lys Ile Thr Pro Arg Phe Gly Ser Glu Val 70 75 80 85							656
cgt ggc ttg caa ctt tcc tcc ttg gat gct gcc gga aag gac gaa ttg Arg Gly Leu Gln Leu Ser Ser Leu Asp Ala Ala Gly Lys Asp Glu Leu 90 95 100							704
gca ctg ttg gct tca caa agg gga gtc tta gtc ttt aga gac cag gac Ala Leu Leu Ala Ser Gln Arg Gly Val Leu Val Phe Arg Asp Gln Asp 105 110 115							752
ttt gca tcc aag ggc ccg ggt tac ttc gag cag tat ggt tcc cat ttc Phe Ala Ser Lys Gly Pro Gly Tyr Phe Glu Gln Tyr Gly Ser His Phe 120 125 130							800
gga aaa tta cat gtt cac caa acc tct ggc gct cct gca ggc cat cca Gly Lys Leu His Val His Gln Thr Ser Gly Ala Pro Ala Gly His Pro 135 140 145							848

tat cta cat gtc aca ttt aga cgc ccc gac aga aag gaa ttt gag agg	896
Tyr Leu His Val Thr Phe Arg Arg Pro Asp Arg Lys Glu Phe Glu Arg	
150 155 160 165	
gtt ttc cgt gac cac cat agc tcc ata cgt ttc cac tcg gat gtg tca	944
Val Phe Arg Asp His His Ser Ser Ile Arg Phe His Ser Asp Val Ser	
170 175 180	
tat gag tta caa cca ccg tct tac aca ttc ttc aca gtt cta gaa ggg	992
Tyr Glu Leu Gln Pro Pro Ser Tyr Thr Phe Phe Thr Val Leu Glu Gly	
185 190 195	
ccc gat ggg gga gga gat act ttg ttt agc gat gcc acg gaa gca ttt	1040
Pro Asp Gly Gly Gly Asp Thr Leu Phe Ser Asp Ala Thr Glu Ala Phe	
200 205 210	
aat agg ctt tcg cct cct ctt cag gac ttt ctg agt aag cta cac gta	1088
Asn Arg Leu Ser Pro Pro Leu Gln Asp Phe Leu Ser Lys Leu His Val	
215 220 225	
ctg cat agc gct aaa gag caa gcc cag aat gcg caa gaa cag ggt gga	1136
Leu His Ser Ala Lys Glu Gln Ala Gln Asn Ala Gln Glu Gln Gly Gly	
230 235 240 245	
att caa cgg aga gaa ccg gtc gca cac att cat cca tta gtg cgt gtc	1184
Ile Gln Arg Arg Glu Pro Val Ala His Ile His Pro Leu Val Arg Val	
250 255 260	
cat cct gtg ctc aag aca aaa agc ttc ttc gtg aat aga ggg ttt aca	1232
His Pro Val Leu Lys Thr Lys Ser Phe Phe Val Asn Arg Gly Phe Thr	
265 270 275	
aga aag att gtg gag ctc aag caa act gaa tcc gat ctg ctg cta gat	1280
Arg Lys Ile Val Glu Leu Lys Gln Thr Glu Ser Asp Leu Leu Leu Asp	
280 285 290	
ttt ata tat aat atc atc gag aac acg cat gat ctc cag tta agg gca	1328
Phe Ile Tyr Asn Ile Ile Glu Asn Thr His Asp Leu Gln Leu Arg Ala	
295 300 305	
agc tgg gaa cct ggt act gta act gtc tgg gac aat agg cgt gtt aac	1376
Ser Trp Glu Pro Gly Thr Val Thr Val Trp Asp Asn Arg Arg Val Asn	
310 315 320 325	
cat tcg gcc atc att gat tgg gaa gag cct gtt tct aga cat ggc gtc	1424
His Ser Ala Ile Ile Asp Trp Glu Glu Pro Val Ser Arg His Gly Val	
330 335 340	
aga atc act cct cat ggc gag aga ccc gtt gag gat ttg aag tat ctg	1472
Arg Ile Thr Pro His Gly Glu Arg Pro Val Glu Asp Leu Lys Tyr Leu	
345 350 355	
aac gat ggg gat tat tac acg agc caa gcg gaa gtg gac tatttcaccg	1521
Asn Asp Gly Asp Tyr Tyr Thr Ser Gln Ala Glu Val Asp	
360 365 370	
aatgagatga tttatgtatt tttgatctgg gcctaaacac atttagatat gaatataggc	1581
tatatatgct tagcctaaac acaatatata catgcatgaa atgtacttgc gtctttttaa	1641
gaaggctaga caatttttaa cactagggaa cgtcaaagaa cttggtaagt ttcaaaacac	1701

```

gcttcatttg gggaaatcttt aagctctctt attgctacat aacgtgattc atggaacttc 1761
ttctaaaaaac ttgaagtttt ctttattaca aagacatata tatatttgta tggcccttat 1821
tatattatga tatatattat taaataaagt actctaaatt tgatccgacg gtcatagaca 1881
tcacatatga atgaattttac aaaaggaggg acagccatcc agcattatta aaaggaagtt 1941
aaatatccaa tttcctcatg agacaacgca ctgatataca cgcatttgct caataagtcc 2001
tcgtcggccc gtatatcata aaaccttgcc cctggactcg ctgtatgaag ttcttcacaa 2061
tttgaacatt agtctggaac ttattgtagt taacaatttc atttgggaca atccagtcgc 2121
gtttct 2127

```

<210> 4

<211> 370

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 70

<400> 4

```

Gly Asn Phe Asp Thr His Phe Phe Ala Glu Gln Asp Glu Val Gly Asp
1           5           10           15

```

```

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

```

```

Pro Lys Phe Leu Pro Thr Trp Asp His Ser Gln Lys Tyr Pro Pro Leu
35           40           45

```

```

Val Phe His Lys Tyr Leu Asp Pro Ala Leu Arg Ala Asp Pro Gln Leu
50           55           60

```

```

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

```

```

Phe Gly Ser Glu Val Arg Gly Leu Gln Leu Ser Ser Leu Asp Ala Ala
85           90           95

```

```

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

```

Phe Arg Asp Gln Asp Phe Ala Ser Lys Gly Pro Gly Tyr Phe Glu Gln
 115 120 125

Tyr Gly Ser His Phe Gly Lys Leu His Val His Gln Thr Ser Gly Ala
 130 135 140

Pro Ala Gly His Pro Tyr Leu His Val Thr Phe Arg Arg Pro Asp Arg
 145 150 155 160

Lys Glu Phe Glu Arg Val Phe Arg Asp His His Ser Ser Ile Arg Phe
 165 170 175

His Ser Asp Val Ser Tyr Glu Leu Gln Pro Pro Ser Tyr Thr Phe Phe
 180 185 190

Thr Val Leu Glu Gly Pro Asp Gly Gly Gly Asp Thr Leu Phe Ser Asp
 195 200 205

Ala Thr Glu Ala Phe Asn Arg Leu Ser Pro Pro Leu Gln Asp Phe Leu
 210 215 220

Ser Lys Leu His Val Leu His Ser Ala Lys Glu Gln Ala Gln Asn Ala
 225 230 235 240

Gln Glu Gln Gly Gly Ile Gln Arg Arg Glu Pro Val Ala His Ile His
 245 250 255

Pro Leu Val Arg Val His Pro Val Leu Lys Thr Lys Ser Phe Phe Val
 260 265 270

Asn Arg Gly Phe Thr Arg Lys Ile Val Glu Leu Lys Gln Thr Glu Ser
 275 280 285

Asp Leu Leu Leu Asp Phe Ile Tyr Asn Ile Ile Glu Asn Thr His Asp
 290 295 300

Leu Gln Leu Arg Ala Ser Trp Glu Pro Gly Thr Val Thr Val Trp Asp
 305 310 315 320

Asn Arg Arg Val Asn His Ser Ala Ile Ile Asp Trp Glu Glu Pro Val
 325 330 335

Ser Arg His Gly Val Arg Ile Thr Pro His Gly Glu Arg Pro Val Glu
 340 345 350

Asp Leu Lys Tyr Leu Asn Asp Gly Asp Tyr Tyr Thr Ser Gln Ala Glu
 355 360 365

Val Asp
 370

<210> 5

<211> 966

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> gene

<222> (399) .. (677)

<223>

<220>

<221> misc_feature

<223> Oligo 74

<400> 5
 gatccgtata gaatgttgaa cttgaaattg tacttttctat ctggaaaatg catctggagg 60
 aagtcagctg tactagtaag actatgaaga ctgaacatga cagaagatat ctcaggatac 120
 taagcgtaga tgtggcacat gcagcaggtg caatatcaaa aaccaaatac agaatttatt 180
 tcacaaaggt acgtaattag ggtctgtag cgcaaaagca cctgctgcat ggcgcagggg 240
 acctcgta caatcataat aaacacagtg agtcaattgt gcgtgtttta cgccatatat 300
 actcatgtaa aaacattcaa cgtctttact atatacctgt ccataactaat ccataactaaa 360
 aaaaccacaa cgaagaaaat tacttggagc tgtcacgccc aatttggatc ttgcggactg 420
 cctgggtccgc gcgaagctct agctttggca cgcgcaacgt gagcacccca tgctcgtagt 480
 tcgcagtgat gtgctctgcg tccacgccag gcttgtccgg caagccaatc acgcgtctga 540
 agctgccagt ggcacgttcc cgcacgcgcc acccattctt gttctcctcg gtctcctttg 600
 taggcacctc acccgatata actagctctc cgcggtcgct gtggaactcg atgttgatgt 660
 tgtcctgggc caccgatgcc ggcacgctca agtgacacct gtagtctttc tcttgctcca 720
 acacatccac gggaggggatg atgctggcag atgtgagcca gtcagtgggtg tctttcttgt 780

tggttggtcctt agcctgcagt gcagtactgt tgggagcgcg accaaggagc ctggtgtagt 840
tggcaacttc gttgttaata ggcgtcaaag aaatcgaaaa atggtgactg gaacgacata 900
ttgctggttag gctgagttgc tatttggtga acactatttc tctttggtga actcttgctt 960
atgatc 966

<210> 6

<211> 11

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 74

<400> 6

Ser Phe Gln Ser Pro Phe Phe Asp Phe Phe Asp
1 5 10

<210> 7

<211> 158

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 74

<400> 7

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

Asn Ser Thr Ala Leu Gln Ala Lys Asn Asn Asn Lys Lys Asp Thr Thr
20 25 30

Asp Trp Leu Thr Ser Ala Ser Ile Ile Pro Pro Val Asp Val Leu Glu
35 40 45

Gln Glu Lys Asp Tyr Glu Val His Leu Ser Val Pro Gly Ile Val Asp
50 55 60

Gln Asp Asn Ile Asn Ile Glu Phe His Ser Asp Arg Gly Glu Leu Val
65 70 75 80

Ile Ser Gly Glu Val Pro Thr Lys Glu Thr Glu Glu Asn Lys Asn Gly
85 90 95

Trp Arg Val Arg Glu Arg Ala Thr Gly Ser Phe Arg Arg Val Ile Gly
100 105 110

Leu Pro Asp Lys Pro Gly Val Asp Ala Glu His Ile Thr Ala Asn Tyr
115 120 125

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

Asp Gln Ala Val Arg Lys Ile Gln Ile Gly Arg Asp Ser Ser
145 150 155

<210> 8

<211> 1260

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> gene

<222> (828) .. (1259)

<223>

<220>

<221> misc_feature

<223> Oligo 113

<220>

<221> misc_feature

<222> (1192)..(1192)

<223> n = unknown nukleotide

<400> 8
gatcccacaa gacccaaggg gataaaaatc aacagaggca cagcaagccc cttagaacga 60
ttagcactta cgatcgcgcc caagatgcac agcatcgaga tcacaagagc tgtagcgcca 120
tagtacggct gcttttctaa ggggaccatt ggtgagaaat ctgcgccatt ctggaactgt 180
ttctgtaatt gagttagtac tccattcttg gagcagtttc ctaaaatcca tgaacataca 240
tagagcttct catatgacat cttgaatctt tcgtgcaact actctcccgt agtatgcaca 300
ctgtatttgt acgtggctaa aaactttttt cccatatgaa ccacccgcac agcggagtgc 360
atcgccactg cagagcgaat gttggctgcc atccgacggc ctctggccac taggcaactg 420
ctgctgagac atgccagtac agaggcatat acgccaccag aataccggga tttaaatgca 480
caaacctggg ccaaattgtc cgagccagtg cgcaagagc tgatagagta tttcgcttgg 540
cgtatggaag accgctggca cacgatatcg cgggatgaag ctctgtccgc ttacttcatc 600
ggatacggga catggggccc acgtgctgcg aagggcaatc caacgcaaca acctcccgcg 660
ggggagcttg tgggtagggc tatattctcg ttgggtctct tctccgcgct aggtgtagca 720
gcattcaatc acgtgaccga caagcgcgtc cagcggcac tcgcgaggct cgaggtgtcc 780
gatgtataag ttatccgtat tatactatat gcaattacac accgctagac attccggagg 840
tagtgggtgg ggcattggta gcccgtaggg ttgttcgtta gatagcgctg gtggtattct 900
tccgcgtcgt agaaagactc agcaagggcc acctcggtct taatgcagtt ggcccagcga 960
ggctgccagc gctcaaggat ctggcgcacc tggctcaagt cctcgtctgg aatgcgcata 1020
caactgcact gctggtactg ggtgcccaca tcagggccct gggcattaag cgtggtgggg 1080
tcgtgcatgc ggaaaaagaa atcacaaagc tcaactcaatc ccaacacctc gggatcgaaa 1140
ggcgacctgg accacctctg cgaaattggt gttacctaca caaactcgct cngtatgtga 1200
ctgcccttcg tgtcggtagg aacattgcc a ttggcgtatc ctgccaccag gtccacgac 1260

<210> 9

<211> 143

<212> PRT

<213> Ashbya gossypii

<220>

<221> MISC_FEATURE

<222> (23)..(23)

<223> X = unknown amino acid

<220>

<221> MISC_FEATURE

<223> Oligo 113

<400> 9

Ile Val Asp Leu Val Ala Gly Tyr Ala Asn Gly Asn Val Pro Thr Asp
1 5 10 15

Thr Lys Gly Ser His Ile Xaa Ser Glu Phe Val Val Thr Pro Ile Ser
20 25 30

Gln Arg Trp Ser Arg Ser Pro Phe Asp Pro Glu Val Leu Gly Leu Ser
35 40 45

Glu Leu Cys Asp Phe Phe Phe Arg Met His Asp Pro Thr Thr Leu Asn
50 55 60

Ala Gln Gly Pro Asp Val Gly Thr Gln Tyr Gln Gln Cys Ser Cys Met
65 70 75 80

Arg Ile Pro Asp Glu Asp Leu Ser Gln Val Arg Gln Ile Leu Glu Arg
85 90 95

Trp Gln Pro Arg Trp Ala Asn Cys Ile Lys Thr Glu Val Ala Leu Ala
100 105 110

Glu Ser Phe Tyr Asp Ala Glu Glu Tyr His Gln Arg Tyr Leu Thr Asn
115 120 125

Asn Pro Thr Gly Tyr Gln Cys Pro Thr His Tyr Leu Arg Asn Val
130 135 140

<210> 10

<211> 2575

<212> DNA

gtg tct ccc tct ctc aaa ttc agt cct caa aaa cac aag ctg ttg aca	824
Val Ser Pro Ser Leu Lys Phe Ser Pro Gln Lys His Lys Leu Leu Thr	
90 95 100	
ttg gct gcc gga tgc ttc tgg gga aca gag cac atg ctt cgt aag cac	872
Leu Ala Ala Gly Cys Phe Trp Gly Thr Glu His Met Leu Arg Lys His	
105 110 115	
ctc ggg cac aag atc gtg gac ctg gtg gca gga tac gcc aat ggc aat	920
Leu Gly His Lys Ile Val Asp Leu Val Ala Gly Tyr Ala Asn Gly Asn	
120 125 130	
gtt cct acc gac acg aag ggc gtc aca tac gag cga gtt tgt gta ggt	968
Val Pro Thr Asp Thr Lys Gly Val Thr Tyr Glu Arg Val Cys Val Gly	
135 140 145 150	
aac acc aat ttc gca gag gtg gtc cag gtc gct ttc gat ccc gag gtg	1016
Asn Thr Asn Phe Ala Glu Val Val Gln Val Ala Phe Asp Pro Glu Val	
155 160 165	
ttg gga ttg agt gag ctt tgt gat ttc ttt ttc cgc atg cac gac ccc	1064
Leu Gly Leu Ser Glu Leu Cys Asp Phe Phe Phe Arg Met His Asp Pro	
170 175 180	
acc acg ctt aat gcc cag ggc cct gat gtg ggc acc cag tac cgc agt	1112
Thr Thr Leu Asn Ala Gln Gly Pro Asp Val Gly Thr Gln Tyr Arg Ser	
185 190 195	
gcg ttg tat gcg cat tcc gac gag gac ttg agc cag gtg cgc cag atc	1160
Ala Leu Tyr Ala His Ser Asp Glu Asp Leu Ser Gln Val Arg Gln Ile	
200 205 210	
ctt gag cgc tgg cag cct cgc tgg gcc aac tgc att aag acc gag gtg	1208
Leu Glu Arg Trp Gln Pro Arg Trp Ala Asn Cys Ile Lys Thr Glu Val	
215 220 225 230	
gcc ctt gct gag tct ttc tac gac gcg gaa gaa tac cac cag cgc tat	1256
Ala Leu Ala Glu Ser Phe Tyr Asp Ala Glu Glu Tyr His Gln Arg Tyr	
235 240 245	
cta acg aac aac ccc acg ggc tac caa tgc ccc acc cac tac ctc cgg	1304
Leu Thr Asn Asn Pro Thr Gly Tyr Gln Cys Pro Thr His Tyr Leu Arg	
250 255 260	
aat gtc tagcgggtgtg taattgcata tagtataata cggataaactt atacatcgga	1360
Asn Val	
cacctcgagc ctcgcgagtg ccgcgtggac gcgcttgctg gtcacgtgat tgaatgctgc	1420
tacacctagc gcggagaaga gcaccaacga gaatatagcc ctaccacaaa gctccccgc	1480
gggaggttgt tgcgttgat tgcccttcgc agcacgtggg ccccatgtcc cgtatccgat	1540
gaagtaagcg gcacgagctt catcccgca tatcgtgtgc cagcggctctt ccatacgcca	1600
ggcgaaatac tctatcagct cttcgcgcac tggctcggac aatttgcccc aggtttgtgc	1660
atttaaatacc cggatattctg gtggcgtata tgccctctgta ctggcatgtc tcagcagcag	1720
ttgcctagtgc gccagaggcc gtcggatggc agccaacatt cgctctgcag tggcgatgaa	1780

```

ctccgctgtg cgggtggttc atatgggaaa aaagttttta gccacgtaca aatacagtgt 1840
gcatactacg ggagagtagt tgcacgaaag attcaagatg tcatatgaga agctctatgt 1900
atgttcatgg attttaggaa actgctccaa gaatggagta ctaactcaat tacagaaaca 1960
gttccagaat ggcgcagatt tctcaccaat ggtcccccta gaaaagcagc cgtactatgg 2020
cgctacagct cttgtgatct cgatgctgtg catcttgggc gcgatcgtaa gtgctaatacg 2080
ttctaagggg cttgctgtgc ctctgttgat ttttatcccc ttgggtcttg tgggatctct 2140
gctgtttggg cttgcaagca tcttcctcgc taactcgttc ggcgtttacg tttgaggttg 2200
cggggacgta tgctagaccg agggcattgc ggggttgagg ggaagagggt gagattatga 2260
atgatataata ctgttatacc ggctgcgggt ggctgtgcgc gtcacacga gggacttaca 2320
agttcaaaaag gtcttcatcg atatttacca acttgtaata acgctcttgt gagtctgagt 2380
tgaggagacc gggctggtcg ccatactcca tcaacgtgtt taccattgcg cgtgtatagc 2440
tgatcatgtt ttcgagggat gactcgtctt tcttcttttag gaacatcaaa atgggtgggtgt 2500
tcacaagctg gaaaagccta tcctgtaggt aagatgggggt gagatacacc gctaattcct 2560
tagggcactc aacca 2575

```

<210> 11

<211> 264

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 113

<400> 11

```

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

```

```

Phe Arg Gln Leu Arg Thr Arg Leu Ala Ala Gly Ala Thr Val Ile Ser
20           25           30

```

```

Arg Asp Phe Leu Phe Arg Leu Val Ile Ser Asn Arg Lys Thr Val Ala
35           40           45

```

```

His Arg Leu Ala Leu Thr Gln Ile Tyr Ile Gln Leu His Asp Thr Leu
50           55           60

```

Ile Arg Thr His Thr Gly Glu His Val Asp Thr Met Thr Phe Asp Thr
65 70 75 80

Pro Pro Pro Ala Ala Pro Val Ser Pro Ser Leu Lys Phe Ser Pro Gln
85 90 95

Lys His Lys Leu Leu Thr Leu Ala Ala Gly Cys Phe Trp Gly Thr Glu
100 105 110

His Met Leu Arg Lys His Leu Gly His Lys Ile Val Asp Leu Val Ala
115 120 125

Gly Tyr Ala Asn Gly Asn Val Pro Thr Asp Thr Lys Gly Val Thr Tyr
130 135 140

Glu Arg Val Cys Val Gly Asn Thr Asn Phe Ala Glu Val Val Gln Val
145 150 155 160

Ala Phe Asp Pro Glu Val Leu Gly Leu Ser Glu Leu Cys Asp Phe Phe
165 170 175

Phe Arg Met His Asp Pro Thr Thr Leu Asn Ala Gln Gly Pro Asp Val
180 185 190

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

Ser Gln Val Arg Gln Ile Leu Glu Arg Trp Gln Pro Arg Trp Ala Asn
210 215 220

Cys Ile Lys Thr Glu Val Ala Leu Ala Glu Ser Phe Tyr Asp Ala Glu
225 230 235 240

Glu Tyr His Gln Arg Tyr Leu Thr Asn Asn Pro Thr Gly Tyr Gln Cys
245 250 255

Pro Thr His Tyr Leu Arg Asn Val
260

<210> 12

<211> 1717

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> gene

<222> (370)..(1608)

<223>

<220>

<221> misc_feature

<223> Oligo 60

<400> 12
gatcccagaa gatgtcgcag ccgccactgg gctctcttta gtctacgaat cacctatagc 60
aaatttggac cttacttatg cactcccttt gaccagcaga ccagtggatt tggctaagcc 120
cgggttttct ttcgggggtcg ccattgcata ccgctgaata catgattata ttataattc 180
aatgtatata attttcggct cttgctgcct ctgccttgat tctatttcag catgtgaggt 240
tcgcttttgt tcaccgattg ctgggactga gacaaaatta atcaacctca agctacggca 300
gaagttcccc ctttccacac atcgcataaa tggggaagag ttggatgtgt atatggaaca 360
caaaggcata tgtgggtttt caaatgtaac tatcaaaaga gcggtcagct tgggtgcactg 420
cagaatatct cctgcgcttt gtgtccagag agagagttta tcgtaggaag agcaagcaaa 480
agtacactag ttatagagga aaaagggtgta tcaagaacac atctccgact agaggtaata 540
gacgaccatg tcaatattct gctattggga tctcagctga aagttggtgg cgaagatgca 600
tctaagggat cgggtgttacc gtttggacca tctcagagcc ctgttattct ggaagtggga 660
tcacattccg tccaatgcat attggagtgg gctcaatggg agttcaaggt cccgaatact 720
ttgttacaag aagaccctgc tttgcttgag gtaattaaag ggtaggaat acgagtgggtg 780
acctcttttt cgaaggccac tacacatcat attattagac cccatgaaga tgaccaagat 840
tacgaaaagt atttgtttgc acttggttaa ggcataccaa tattgcatct cgaatttctg 900
aaacaattta ccgctcttct tcgacagcct gttaccgatt ttgaccaca gctacgcagt 960
atacaagagc agcattacgc ctttagaggc ttactaaga aagccgcaga cctgaatggg 1020
ttgcacttta tagtaaacag aaaacatgat tttgaaatgc taaaatatac attggagata 1080
ggaggtggaa cgggtgcattt ttgcatgac attaatactg tgagcgcagt tctcaagaca 1140
atacctatta atggagctct tgcattgaaa tatgatcatg ggatatctgc agaggtaaag 1200
aaagttgacg atacacgata ttgtgaggtc ttgcactcgc atggtatata gttgtttact 1260

```

gtccatgaac taacgaaagc catactgtac gataagcttg ctcagttgct aagtaagagg 1320
ccaccacta tgatgaaatt agtggaaacca gtgccatcta ctatatctca aacacaagaa 1380
atgaaacaag aatctttatc acccatctca ttgggaaata cgactatcaa ggatttaccg 1440
cgtaaagaca aaacaaggaa acgggtcaag aaacctcagc ctctggatag tttgtcattt 1500
tttgcaggtg gtctatcgcg gcagaatagc tcgataggtg gtacagctga ccttggtaaa 1560
caagaaagca tgagcttagg agcagatact cattatggaa gtccaaacga tgtccaatcg 1620
ttaccgcga cccaaagccc gaaaaagagg aggcgtccaa aggttgaacc gctagggaat 1680
ttaatgacct ctcaattgca aaaagctttc gatgatc 1717

```

<210> 13

<211> 413

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 60

<400> 13

```

Met Trp Val Phe Lys Cys Asn Tyr Gln Lys Ser Gly Gln Leu Gly Ala
1           5           10          15

```

```

Leu Gln Asn Ile Ser Cys Ala Leu Cys Pro Glu Arg Glu Phe Ile Val
20          25          30

```

```

Gly Arg Ala Ser Lys Ser Thr Leu Val Ile Glu Glu Lys Gly Val Ser
35          40          45

```

```

Arg Thr His Leu Arg Leu Glu Val Ile Asp Asp His Val Asn Ile Leu
50          55          60

```

```

Leu Leu Gly Ser Gln Leu Lys Val Gly Gly Glu Asp Ala Ser Lys Gly
65          70          75          80

```

```

Ser Val Leu Pro Phe Gly Pro Ser Gln Ser Pro Val Ile Leu Glu Val
85          90          95

```

Gly Ser His Ser Val Gln Cys Ile Leu Glu Trp Ala Gln Trp Glu Phe
 100 105 110

Lys Val Pro Asn Thr Leu Leu Gln Glu Asp Pro Ala Leu Leu Glu Val
 115 120 125

Ile Lys Gly Leu Gly Ile Arg Val Val Thr Ser Phe Ser Lys Ala Thr
 130 135 140

Thr His His Ile Ile Arg Pro His Glu Asp Asp Gln Asp Tyr Glu Lys
 145 150 155 160

Tyr Leu Phe Ala Leu Val Lys Gly Ile Pro Ile Leu His Leu Glu Phe
 165 170 175

Leu Lys Gln Phe Thr Ala Leu Leu Arg Gln Pro Val Thr Asp Phe Asp
 180 185 190

Pro Gln Leu Arg Ser Ile Gln Glu Gln His Tyr Ala Phe Arg Gly Phe
 195 200 205

Thr Lys Lys Ala Ala Asp Leu Asn Gly Leu His Phe Ile Val Asn Arg
 210 215 220

Lys His Asp Phe Glu Met Leu Lys Tyr Thr Leu Glu Ile Gly Gly Gly
 225 230 235 240

Thr Val His Phe Cys His Asp Ile Asn Asp Leu Ser Ala Cys Leu Lys
 245 250 255

Thr Ile Pro Ile Asn Gly Ala Leu Ala Leu Lys Tyr Asp His Gly Ile
 260 265 270

Ser Ala Glu Val Lys Lys Val Asp Asp Thr Arg Tyr Cys Glu Val Leu
 275 280 285

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

Ile Leu Tyr Asp Lys Leu Ala Gln Leu Leu Ser Lys Arg Pro Pro Thr
 305 310 315 320

Met Met Lys Leu Val Glu Pro Val Pro Ser Thr Ile Ser Gln Thr Gln
 325 330 335

Glu Met Lys Gln Glu Ser Leu Ser Pro Ile Ser Leu Gly Asn Thr Thr
 340 345 350

Ile Lys Asp Leu Pro Arg Lys Asp Lys Thr Arg Lys Arg Leu Lys Lys
 355 360 365

Pro Gln Pro Leu Asp Ser Leu Ser Phe Phe Ala Gly Gly Leu Ser Arg
 370 375 380

Gln Asn Ser Ser Ile Gly Gly Thr Ala Asp Leu Val Lys Gln Glu Ser
 385 390 395 400

Met Ser Leu Gly Ala Asp Thr His Tyr Gly Ser Pro Asn
 405 410

<210> 14

<211> 3746

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (570) .. (3176)

<223>

<220>

<221> misc_feature

<223> Oligo 60

<400> 14
 catatcatgg accaatttta cctgggcggg ctgtcatctt tgaagggatt tcatcggaac 60
 tctgtcgga gtagaggagg tgacttatcc tacagacttt ccttaaagtc gtctgtccac 120
 ttaccaggca cgccgctgga ctgcgctttg aggctacaag ccttcttcaa tgccggtgat 180
 gttttccact ttaggcgcca gatcccagaa gatgtcgcag ccgccactgg gctctcttta 240
 gtctacgaat cacctatagc aaatttggac cttacttatg cactcccttt gaccagcaga 300
 ccagtggatt tggctaagcc cgggttttct ttcggggtcg ccattgcata ccgctgaata 360
 catgattata ttataaatcc aatgtatata attttcggtt cttgctgcct ctgccttgat 420
 tctatttcag catgtgaggt tcgcttttgt tcaccgattg ctcggaactga gacaaaatta 480

atcaacctca agctacggca gaagttcccc ctttccacac atcgcataaa tggggaagag 540

ttggatgtgt atatggaaca caaaggcat atg tgg gtt ttc aaa tgt aac tat 593
Met Trp Val Phe Lys Cys Asn Tyr
1 5

caa aag agc ggt cag ctt ggt gca ctg cag aat atc tcc tgc gct ttg 641
Gln Lys Ser Gly Gln Leu Gly Ala Leu Gln Asn Ile Ser Cys Ala Leu
10 15 20

tgt cca gag aga gag ttt atc gta gga aga gca agc aaa agt aca cta 689
Cys Pro Glu Arg Glu Phe Ile Val Gly Arg Ala Ser Lys Ser Thr Leu
25 30 35 40

gtt ata gag gaa aaa ggt gta tca aga aca cat ctc cga cta gag gta 737
Val Ile Glu Glu Lys Gly Val Ser Arg Thr His Leu Arg Leu Glu Val
45 50 55

ata gac gac cat gtc aat att ctg cta ttg gga tct cag ctg aaa gtt 785
Ile Asp Asp His Val Asn Ile Leu Leu Leu Gly Ser Gln Leu Lys Val
60 65 70

ggt ggc gaa gat gca tct aag gga tcg gtg tta ccg ttt gga cca tct 833
Gly Gly Glu Asp Ala Ser Lys Gly Ser Val Leu Pro Phe Gly Pro Ser
75 80 85

cag agc cct gtt att ctg gaa gtg gga tca cat tcc gtc caa tgc ata 881
Gln Ser Pro Val Ile Leu Glu Val Gly Ser His Ser Val Gln Cys Ile
90 95 100

ttg gag tgg gct caa tgg gag ttc aag gtc ccg aat act ttg tta caa 929
Leu Glu Trp Ala Gln Trp Glu Phe Lys Val Pro Asn Thr Leu Leu Gln
105 110 115 120

gaa gac cct gct ttg ctt gag gta att aaa ggg tta gga ata cga gtg 977
Glu Asp Pro Ala Leu Leu Glu Val Ile Lys Gly Leu Gly Ile Arg Val
125 130 135

gtg acc tot ttt tcg aag gcc act aca cat cat att att aga ccc cat 1025
Val Thr Ser Phe Ser Lys Ala Thr Thr His His Ile Ile Arg Pro His
140 145 150

gaa gat gac caa gat tac gaa aag tat ttg ttt gca ctt gtt aaa ggc 1073
Glu Asp Asp Gln Asp Tyr Glu Lys Tyr Leu Phe Ala Leu Val Lys Gly
155 160 165

ata cca ata ttg cat ctc gaa ttt ctg aaa caa ttt acc gct ctt ctt 1121
Ile Pro Ile Leu His Leu Glu Phe Leu Lys Gln Phe Thr Ala Leu Leu
170 175 180

cga cag cct gtt acc gat ttt gac cca cag cta cgc agt ata caa gag 1169
Arg Gln Pro Val Thr Asp Phe Asp Pro Gln Leu Arg Ser Ile Gln Glu
185 190 195 200

cag cat tac gcc ttt aga ggc ttt act aag aaa gcc gca gac ctg aat 1217
Gln His Tyr Ala Phe Arg Gly Phe Thr Lys Lys Ala Ala Asp Leu Asn
205 210 215

ggg ttg cac ttt ata gta aac aga aaa cat gat ttt gaa atg cta aaa 1265
Gly Leu His Phe Ile Val Asn Arg Lys His Asp Phe Glu Met Leu Lys
220 225 230

tat aca ttg gag ata gga ggt gga acg gtg cat ttt tgc cat gac att Tyr Thr Leu Glu Ile Gly Gly Gly Thr Val His Phe Cys His Asp Ile 235 240 245	1313
aat gac ttg agc gca tgt ctc aag aca ata cct att aat gga gct ctt Asn Asp Leu Ser Ala Cys Leu Lys Thr Ile Pro Ile Asn Gly Ala Leu 250 255 260	1361
gca ttg aaa tat gat cat ggg ata tct gca gag gta aag aaa gtt gac Ala Leu Lys Tyr Asp His Gly Ile Ser Ala Glu Val Lys Lys Val Asp 265 270 275 280	1409
gat aca cga tat tgt gag gtc ttg cac tcg cat ggt ata cag ttg ttt Asp Thr Arg Tyr Cys Glu Val Leu His Ser His Gly Ile Gln Leu Phe 285 290 295	1457
act gtc cat gaa cta acg aaa gcc ata ctg tac gat aag ctt gct cag Thr Val His Glu Leu Thr Lys Ala Ile Leu Tyr Asp Lys Leu Ala Gln 300 305 310	1505
ttg cta agt aag agg cca ccc act atg atg aaa tta gtg gaa cca gtg Leu Leu Ser Lys Arg Pro Pro Thr Met Met Lys Leu Val Glu Pro Val 315 320 325	1553
cca tct act ata tct caa aca caa gaa atg aaa caa gaa tct tta tca Pro Ser Thr Ile Ser Gln Thr Gln Glu Met Lys Gln Glu Ser Leu Ser 330 335 340	1601
ccc atc tca ttg gga aat acg act atc aag gat tta ccg cgt aaa gac Pro Ile Ser Leu Gly Asn Thr Thr Ile Lys Asp Leu Pro Arg Lys Asp 345 350 355 360	1649
aaa aca agg aaa cgg ctc aag aaa cct cag cct ctg gat agt ttg tca Lys Thr Arg Lys Arg Leu Lys Lys Pro Gln Pro Leu Asp Ser Leu Ser 365 370 375	1697
ttt ttt gca ggt ggt cta tcg cgg cag aat agc tcg ata ggt ggt aca Phe Phe Ala Gly Gly Leu Ser Arg Gln Asn Ser Ser Ile Gly Gly Thr 380 385 390	1745
gct gac ctt gtt aaa caa gaa agc atg agc tta gga gca gat act cat Ala Asp Leu Val Lys Gln Glu Ser Met Ser Leu Gly Ala Asp Thr His 395 400 405	1793
tat gga agt cca aac gat gtc caa tcg tta ccc gcg acc caa agc ccg Tyr Gly Ser Pro Asn Asp Val Gln Ser Leu Pro Ala Thr Gln Ser Pro 410 415 420	1841
aaa aag agg agg cgt cca aag gtt gaa ccg cta ggg aat tta atg acc Lys Lys Arg Arg Arg Pro Lys Val Glu Pro Leu Gly Asn Leu Met Thr 425 430 435 440	1889
tct caa ttg caa aaa gct ttc gat gat cca act att tcg agc gcg aaa Ser Gln Leu Gln Lys Ala Phe Asp Asp Pro Thr Ile Ser Ser Ala Lys 445 450 455	1937
gta tcg ccc act gag cag gaa att atc cca gtt cgt gaa tca caa gct Val Ser Pro Thr Glu Gln Glu Ile Ile Pro Val Arg Glu Ser Gln Ala 460 465 470	1985
aaa aaa acc ccc gac ccc att ggg gca ttt aat ggc act tca cag aag	2033

Lys	Lys	Thr	Pro	Asp	Pro	Ile	Gly	Ala	Phe	Asn	Gly	Thr	Ser	Gln	Lys	
	475						480					485				
ggg	agt	ata	gaa	tcg	aac	tca	agg	aag	agg	cag	ccg	agc	cct	gga	tat	2081
Gly	Ser	Ile	Glu	Ser	Asn	Ser	Arg	Lys	Arg	Gln	Pro	Ser	Pro	Gly	Tyr	
	490					495					500					
ccc	tac	gaa	ccc	cta	ccg	gcc	gga	gta	ggg	aat	gga	gat	atc	ctc	aat	2129
Pro	Tyr	Glu	Pro	Leu	Pro	Ala	Gly	Val	Gly	Asn	Gly	Asp	Ile	Leu	Asn	
505					510					515				520		
aag	caa	tgg	aaa	tcc	tta	gat	gag	cag	cca	cct	tct	aaa	aaa	cct	aga	2177
Lys	Gln	Trp	Lys	Ser	Leu	Asp	Glu	Gln	Pro	Pro	Ser	Lys	Lys	Pro	Arg	
				525					530					535		
atg	caa	gaa	aca	gac	ctg	tac	tcg	gaa	gtg	atg	tcg	gga	agt	agc	gaa	2225
Met	Gln	Glu	Thr	Asp	Leu	Tyr	Ser	Glu	Val	Met	Ser	Gly	Ser	Ser	Glu	
			540					545					550			
gag	tat	cgg	gag	gaa	cta	acc	aga	tca	aaa	ctt	aat	tcc	tta	gat	gag	2273
Glu	Tyr	Arg	Glu	Glu	Leu	Thr	Arg	Ser	Lys	Leu	Asn	Ser	Leu	Asp	Glu	
		555					560					565				
gaa	tcc	att	aag	agc	cga	aag	aga	aca	agt	tca	gaa	aag	cct	gtg	aat	2321
Glu	Ser	Ile	Lys	Ser	Arg	Lys	Arg	Thr	Ser	Ser	Glu	Lys	Pro	Val	Asn	
	570					575					580					
agt	tgt	aaa	cgt	cgg	cag	gct	cct	gct	gta	ccg	gaa	caa	aac	ata	gtt	2369
Ser	Cys	Lys	Arg	Arg	Gln	Ala	Pro	Ala	Val	Pro	Glu	Gln	Asn	Ile	Val	
585					590					595					600	
gat	gct	atc	aaa	caa	ata	aag	gaa	cgt	gaa	gtc	aac	cgc	ata	aga	acc	2417
Asp	Ala	Ile	Lys	Gln	Ile	Lys	Glu	Arg	Glu	Val	Asn	Arg	Ile	Arg	Thr	
				605					610					615		
aca	ata	gta	gaa	ctg	ggc	ccc	gag	gaa	tta	acc	gaa	gat	gct	ata	agg	2465
Thr	Ile	Val	Glu	Leu	Gly	Pro	Glu	Glu	Leu	Thr	Glu	Asp	Ala	Ile	Arg	
			620					625					630			
caa	ctt	aaa	tcg	tta	gcg	att	gtc	cag	cca	gta	gat	atg	ctc	agg	caa	2513
Gln	Leu	Lys	Ser	Leu	Ala	Ile	Val	Gln	Pro	Val	Asp	Met	Leu	Arg	Gln	
		635					640					645				
caa	cat	act	gta	gat	aca	gtt	ggg	tcg	tcg	cca	tta	acc	ttt	gca	cgg	2561
Gln	His	Thr	Val	Asp	Thr	Val	Gly	Ser	Ser	Pro	Leu	Thr	Phe	Ala	Arg	
	650					655					660					
aag	aac	tac	aag	aat	ttt	aag	aag	gtg	tgg	ccc	aaa	tat	atg	aat	cgc	2609
Lys	Asn	Tyr	Lys	Asn	Phe	Lys	Lys	Val	Trp	Pro	Lys	Tyr	Met	Asn	Arg	
665					670					675				680		
aag	aat	tcg	cct	gaa	aca	agg	aat	gct	acc	gcg	atc	agt	aac	aga	aaa	2657
Lys	Asn	Ser	Pro	Glu	Thr	Arg	Asn	Ala	Thr	Ala	Ile	Ser	Asn	Arg	Lys	
				685					690					695		
tat	ctc	cca	ttg	gaa	ctt	tat	gaa	act	gat	ctt	cag	agc	cgg	acg	tta	2705
Tyr	Leu	Pro	Leu	Glu	Leu	Tyr	Glu	Thr	Asp	Leu	Gln	Ser	Arg	Thr	Leu	
			700					705					710			
ttt	gaa	gat	ggg	gcg	cct	cca	cga	cag	aag	gac	aat	att	tct	cgt	tct	2753
Phe	Glu	Asp	Gly	Ala	Pro	Pro	Arg	Gln	Lys	Asp	Asn	Ile	Ser	Arg	Ser	
		715					720					725				

act tct aca agt ggc ctc tcg aac gac ctt gca agt aac cgc gga aag Thr Ser Thr Ser Gly Leu Ser Asn Asp Leu Ala Ser Asn Arg Gly Lys 730 735 740	2801
gac cat ccg gct cct gga agt tta ttt tcg aac cat gca tta cta gga Asp His Pro Ala Pro Gly Ser Leu Phe Ser Asn His Ala Leu Leu Gly 745 750 755 760	2849
aat gat aac gca ggc cct gag cca atc ttc aat ttt agt tcc cgt tcc Asn Asp Asn Ala Gly Pro Glu Pro Ile Phe Asn Phe Ser Ser Arg Ser 765 770 775	2897
cgc gag gac aca aat tct cct aca gcg tca cca gct cca aac ctg aca Arg Glu Asp Thr Asn Ser Pro Thr Ala Ser Pro Ala Pro Asn Leu Thr 780 785 790	2945
cat gat aac act tta ttt gtg gta gat gac gaa aac gaa gat ggc gac His Asp Asn Thr Leu Phe Val Val Asp Asp Glu Asn Glu Asp Gly Asp 795 800 805	2993
gga atg gct gat aca agc att gcc cag aga gtc tca agg gag aga cat Gly Met Ala Asp Thr Ser Ile Ala Gln Arg Val Ser Arg Glu Arg His 810 815 820	3041
gct act act aaa cct gcg gaa ccc gat acg gct gag cca cga cac act Ala Thr Thr Lys Pro Ala Glu Pro Asp Thr Ala Glu Pro Arg His Thr 825 830 835 840	3089
cgt cga aaa cat gct ccc aga cca tcc gcg gcc gca gca gac agc gat Arg Arg Lys His Ala Pro Arg Pro Ser Ala Ala Ala Ala Asp Ser Asp 845 850 855	3137
gaa gac ggc gaa cca acg ttc gca ttt aaa cgc gca cga tgaacatttt Glu Asp Gly Glu Pro Thr Phe Ala Phe Lys Arg Ala Arg 860 865	3186
acacataaaa gtttatatag tatattattg acttactcaa gctatcagtt gccaaactac	3246
gggaccaagt cgcgccccgc ccgtctattg gaagagtggg aagggaactaa aatctggatc	3306
aacaaatctt tgaacacctt cctttttgtg aatgtcgtta agcagtttga tatcttcctc	3366
atccattttg aagtattcaa agttagagat gatacggctc tctttcacag acttaggcaa	3426
cacaacaata ccacggttcg caacgtaatt tagcaagacg tgagctggag tgaccccgtg	3486
cttttccgca accttggtca cagtagcgtc ctttagcaat ggagcagagg atgagcccaa	3546
aggcgaatat gcctcaacaa cgataccctt cttctgacaa tactccacaa gctccaactg	3606
aggcaaacgg atatgtagct ccacctggtt gacggctggg acgactgtgg tgccctcggc	3666
attcagtagc tcttctaaat tggtgatgga gaagttcgaa acaccaatgg ccttgaactt	3726
tcccgggtgcc tagcaacttc	3746
<210> 15	
<211> 869	

<212> PRT

<213> *Ashbya gossypii*

<220>

<221> misc_feature

<223> Oligo 60

<400> 15

Met Trp Val Phe Lys Cys Asn Tyr Gln Lys Ser Gly Gln Leu Gly Ala
 1 5 10 15

Leu Gln Asn Ile Ser Cys Ala Leu Cys Pro Glu Arg Glu Phe Ile Val
 20 25 30

Gly Arg Ala Ser Lys Ser Thr Leu Val Ile Glu Glu Lys Gly Val Ser
 35 40 45

Arg Thr His Leu Arg Leu Glu Val Ile Asp Asp His Val Asn Ile Leu
 50 55 60

Leu Leu Gly Ser Gln Leu Lys Val Gly Gly Glu Asp Ala Ser Lys Gly
 65 70 75 80

Ser Val Leu Pro Phe Gly Pro Ser Gln Ser Pro Val Ile Leu Glu Val
 85 90 95

Gly Ser His Ser Val Gln Cys Ile Leu Glu Trp Ala Gln Trp Glu Phe
 100 105 110

Lys Val Pro Asn Thr Leu Leu Gln Glu Asp Pro Ala Leu Leu Glu Val
 115 120 125

Ile Lys Gly Leu Gly Ile Arg Val Val Thr Ser Phe Ser Lys Ala Thr
 130 135 140

Thr His His Ile Ile Arg Pro His Glu Asp Asp Gln Asp Tyr Glu Lys
 145 150 155 160

Tyr Leu Phe Ala Leu Val Lys Gly Ile Pro Ile Leu His Leu Glu Phe
 165 170 175

Leu Lys Gln Phe Thr Ala Leu Leu Arg Gln Pro Val Thr Asp Phe Asp
 180 185 190

Pro Gln Leu Arg Ser Ile Gln Glu Gln His Tyr Ala Phe Arg Gly Phe
 195 200 205

Thr Lys Lys Ala Ala Asp Leu Asn Gly Leu His Phe Ile Val Asn Arg
 210 215 220

Lys His Asp Phe Glu Met Leu Lys Tyr Thr Leu Glu Ile Gly Gly Gly
 225 230 235 240

Thr Val His Phe Cys His Asp Ile Asn Asp Leu Ser Ala Cys Leu Lys
 245 250 255

Thr Ile Pro Ile Asn Gly Ala Leu Ala Leu Lys Tyr Asp His Gly Ile
 260 265 270

Ser Ala Glu Val Lys Lys Val Asp Asp Thr Arg Tyr Cys Glu Val Leu
 275 280 285

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

Ile Leu Tyr Asp Lys Leu Ala Gln Leu Leu Ser Lys Arg Pro Pro Thr
 305 310 315 320

Met Met Lys Leu Val Glu Pro Val Pro Ser Thr Ile Ser Gln Thr Gln
 325 330 335

Glu Met Lys Gln Glu Ser Leu Ser Pro Ile Ser Leu Gly Asn Thr Thr
 340 345 350

Ile Lys Asp Leu Pro Arg Lys Asp Lys Thr Arg Lys Arg Leu Lys Lys
 355 360 365

Pro Gln Pro Leu Asp Ser Leu Ser Phe Phe Ala Gly Gly Leu Ser Arg
 370 375 380

Gln Asn Ser Ser Ile Gly Gly Thr Ala Asp Leu Val Lys Gln Glu Ser
 385 390 395 400

Met Ser Leu Gly Ala Asp Thr His Tyr Gly Ser Pro Asn Asp Val Gln
 405 410 415

Ser Leu Pro Ala Thr Gln Ser Pro Lys Lys Arg Arg Arg Pro Lys Val
 420 425 430

Glu Pro Leu Gly Asn Leu Met Thr Ser Gln Leu Gln Lys Ala Phe Asp
 435 440 445

Asp Pro Thr Ile Ser Ser Ala Lys Val Ser Pro Thr Glu Gln Glu Ile
 450 455 460

Ile Pro Val Arg Glu Ser Gln Ala Lys Lys Thr Pro Asp Pro Ile Gly
 465 470 475 480

Ala Phe Asn Gly Thr Ser Gln Lys Gly Ser Ile Glu Ser Asn Ser Arg
 485 490 495

Lys Arg Gln Pro Ser Pro Gly Tyr Pro Tyr Glu Pro Leu Pro Ala Gly
 500 505 510

Val Gly Asn Gly Asp Ile Leu Asn Lys Gln Trp Lys Ser Leu Asp Glu
 515 520 525

Gln Pro Pro Ser Lys Lys Pro Arg Met Gln Glu Thr Asp Leu Tyr Ser
 530 535 540

Glu Val Met Ser Gly Ser Ser Glu Glu Tyr Arg Glu Glu Leu Thr Arg
 545 550 555 560

Ser Lys Leu Asn Ser Leu Asp Glu Glu Ser Ile Lys Ser Arg Lys Arg
 565 570 575

Thr Ser Ser Glu Lys Pro Val Asn Ser Cys Lys Arg Arg Gln Ala Pro
 580 585 590

Ala Val Pro Glu Gln Asn Ile Val Asp Ala Ile Lys Gln Ile Lys Glu
 595 600 605

Arg Glu Val Asn Arg Ile Arg Thr Thr Ile Val Glu Leu Gly Pro Glu
 610 615 620

Glu Leu Thr Glu Asp Ala Ile Arg Gln Leu Lys Ser Leu Ala Ile Val
 625 630 635 640

Gln Pro Val Asp Met Leu Arg Gln Gln His Thr Val Asp Thr Val Gly
 645 650 655

Ser Ser Pro Leu Thr Phe Ala Arg Lys Asn Tyr Lys Asn Phe Lys Lys
 660 665 670

Val Trp Pro Lys Tyr Met Asn Arg Lys Asn Ser Pro Glu Thr Arg Asn
 675 680 685

Ala Thr Ala Ile Ser Asn Arg Lys Tyr Leu Pro Leu Glu Leu Tyr Glu
 690 695 700

Thr Asp Leu Gln Ser Arg Thr Leu Phe Glu Asp Gly Ala Pro Pro Arg
 705 710 715 720

Gln Lys Asp Asn Ile Ser Arg Ser Thr Ser Thr Ser Gly Leu Ser Asn
 725 730 735

Asp Leu Ala Ser Asn Arg Gly Lys Asp His Pro Ala Pro Gly Ser Leu
 740 745 750

Phe Ser Asn His Ala Leu Leu Gly Asn Asp Asn Ala Gly Pro Glu Pro
 755 760 765

Ile Phe Asn Phe Ser Ser Arg Ser Arg Glu Asp Thr Asn Ser Pro Thr
 770 775 780

Ala Ser Pro Ala Pro Asn Leu Thr His Asp Asn Thr Leu Phe Val Val
 785 790 795 800

Asp Asp Glu Asn Glu Asp Gly Asp Gly Met Ala Asp Thr Ser Ile Ala
 805 810 815

Gln Arg Val Ser Arg Glu Arg His Ala Thr Thr Lys Pro Ala Glu Pro
 820 825 830

Asp Thr Ala Glu Pro Arg His Thr Arg Arg Lys His Ala Pro Arg Pro
 835 840 845

Ser Ala Ala Ala Ala Asp Ser Asp Glu Asp Gly Glu Pro Thr Phe Ala
 850 855 860

Phe Lys Arg Ala Arg
 865

<210> 16

<211> 558

<212> DNA

<213> Ashbya gossypii

<220>

<221> gene

<222> (193) .. (558)

<223>

<220>

<221> misc_feature

<223> Oligo 161

<400> 16
 gatctcctct ccagaagatc cgctaaggta aatacagaat gcgcagttgt ataacagagc 60
 tgtgtgaacg cccgtgagga cgcargaaga gccgcttcgt gcagtcgtat gctactacgg 120
 aggaatttca gaagttcaaa gccaaaaggc cagaatggta gaaacagata tgtacatata 180
 tattccgaat cagtgagcgt ggggatcctc ctttggaggt tctggatcga ccgtaatcat 240
 tgtgtcaatt ctaagaatgg caatgcaagc ttcgagggcg gacttcaaag actttttctt 300
 gctgatggtg ggttccaaca cgcgggctcc cacctcgtcg acgaccttcc cgcggataag 360
 gtcgagcccg tagttgcggg acttmcgctt cttgaggtcg tccggtttgg cctgctggct 420
 ggctgcatgg tacgaacgca gcttggcaac gaggtcgctg gagtccttgg ctgcattgac 480
 ggccaatgtc tttgggatga ccaacatcgc ggcggaacac tcggcgattg ccagctgctc 540
 gcgcgagccg acggtggt 558

<210> 17

<211> 122

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 161

<400> 17

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

Ala Met Leu Val Ile Pro Lys Thr Leu Ala Val Asn Ala Ala Lys Asp
 20 25 30

Ser Ser Asp Leu Val Ala Lys Leu Arg Ser Tyr His Ala Ala Ser Gln
 35 40 45

Gln Ala Lys Pro Asp Asp Leu Lys Lys Arg Lys Tyr Arg Asn Tyr Gly
 50 55 60

Leu Asp Leu Ile Arg Gly Lys Val Val Asp Glu Val Gly Ala Gly Val
 65 70 75 80

Leu Glu Pro Thr Ile Ser Lys Lys Lys Ser Leu Lys Ser Ala Leu Glu
 85 90 95

Ala Cys Ile Ala Ile Leu Arg Ile Asp Thr Met Ile Thr Val Asp Pro
 100 105 110

Glu Pro Pro Lys Glu Asp Pro His Ala His
 115 120

<210> 18

<211> 2341

<212> DNA

<213> *Ashbya gossypii*

<220>

<221> CDS

<222> (479) .. (2152)

<223>

<220>

<221> misc_feature

<223> Oligo 161

<400> 18
 tggccttttgg atacttaagt actaaagaca catagggtcaa tgattcatgt ttttctattg 60
 tatcacgccca ccggcctgct cgcgcccaca gcgtgacccc gaccaagtgc cagctctagc 120
 gccaggaccc gcgcgagacc ttgcgccgcg acaccgcttg cgctgctgca tcgccttata 180
 actgctatat ctgtctttgc gccctgcca aaaatcggca attttggtc tgcacaaaat 240

ttgtggatag agcagcaaag acaccggaag agtcagccgt cgcagactct tgccgccaag	300
acgtatatac ggttggatat caatagtgcg atcacgtggc cactgcgcac gtgacgatac	360
atagtcatta tgaacgaaaa tggcttgcaa agaagagacc gggtaactgg tgggcacggg	420
acttgagcgc cctaaggtgg caaaccgata acaacaagag catcgaagct aaagcagg	478
atg gct caa gtg ttc aac aac cag cgt tcg gat acc ctg ttt ctg ggc	526
Met Ala Gln Val Phe Asn Asn Gln Arg Ser Asp Thr Leu Phe Leu Gly	
1 5 10 15	
ggc gag aag gtg tct gga gat gac atc cgc aac cag aat gtg ctt gcg	574
Gly Glu Lys Val Ser Gly Asp Asp Ile Arg Asn Gln Asn Val Leu Ala	
20 25 30	
gcg ctg gca gtg gcc aat gtg gtg aag tcg tcg ctg ggc ccg gtg ggg	622
Ala Leu Ala Val Ala Asn Val Val Lys Ser Ser Leu Gly Pro Val Gly	
35 40 45	
ctg gac aag atg ctg gtg gac gac atc ggg gac ttc acg gtg acg aac	670
Leu Asp Lys Met Leu Val Asp Asp Ile Gly Asp Phe Thr Val Thr Asn	
50 55 60	
gac ggg gcg acg atc ctg tcg ctg ctg gac gtg cag cac cct gcg ggc	718
Asp Gly Ala Thr Ile Leu Ser Leu Leu Asp Val Gln His Pro Ala Gly	
65 70 75 80	
aag atc ctg gta gag ctg gcg cag cag cag gac cgc gag atc ggc gat	766
Lys Ile Leu Val Glu Leu Ala Gln Gln Gln Asp Arg Glu Ile Gly Asp	
85 90 95	
ggc acg acg agc gtg gtg atc atc gcg gca gag ctg ctg aag cgc gcg	814
Gly Thr Thr Ser Val Val Ile Ile Ala Ala Glu Leu Leu Lys Arg Ala	
100 105 110	
aac gag ctg gtc aag aac aag atc cac ccg acg acg att atc aca ggc	862
Asn Glu Leu Val Lys Asn Lys Ile His Pro Thr Thr Ile Ile Thr Gly	
115 120 125	
ttc cgg gtg gcg ctg cgg gag gcg ctg cgc tac att gaa gac gtg cta	910
Phe Arg Val Ala Leu Arg Glu Ala Leu Arg Tyr Ile Glu Asp Val Leu	
130 135 140	
tcg cag cca gtg gac gcc ctg gcg cgc gag acg atg gtg aac atc gcg	958
Ser Gln Pro Val Asp Ala Leu Ala Arg Glu Thr Met Val Asn Ile Ala	
145 150 155 160	
aag acc agt atg tcg tcc aag atc att ggc tcc gac tcg gag ttc ttt	1006
Lys Thr Ser Met Ser Ser Lys Ile Ile Gly Ser Asp Ser Glu Phe Phe	
165 170 175	
agc tcc atg gtg gtg gat gcg ctg ctg gcc gtc aag acg cag aac ggc	1054
Ser Ser Met Val Val Asp Ala Leu Leu Ala Val Lys Thr Gln Asn Gly	
180 185 190	
cgc ggt gag gta aag tac ccg gtc aag gcg gtg aac ata ctg aag gcg	1102
Arg Gly Glu Val Lys Tyr Pro Val Lys Ala Val Asn Ile Leu Lys Ala	
195 200 205	

cac ggt aaa tcc gcg tgc gag tcc gtg ctc gtg cag ggc tac gcc cta	1150
His Gly Lys Ser Ala Cys Glu Ser Val Leu Val Gln Gly Tyr Ala Leu	
210 215 220	
aac tgt act gtt gca tcg cag gcc atg ccc cgt aaa ctg acg ggc agt	1198
Asn Cys Thr Val Ala Ser Gln Ala Met Pro Arg Lys Leu Thr Gly Ser	
225 230 235 240	
gtg aag atc gcg tgt ctt gat atc aac ctg cag aag gca cgc atg gcg	1246
Val Lys Ile Ala Cys Leu Asp Ile Asn Leu Gln Lys Ala Arg Met Ala	
245 250 255	
atg ggt gtt cag atc aat gtc gac gat ccg gac cag ctg gag gag atc	1294
Met Gly Val Gln Ile Asn Val Asp Asp Pro Asp Gln Leu Glu Glu Ile	
260 265 270	
cgg aag cgg gag gtt gga ata gtg ttg gag cgt gtc aag aaa att att	1342
Arg Lys Arg Glu Val Gly Ile Val Leu Glu Arg Val Lys Lys Ile Ile	
275 280 285	
gac gcc ggg gcc cag gtt gtg ctc aca act aag ggt atc gac gac cta	1390
Asp Ala Gly Ala Gln Val Val Leu Thr Thr Lys Gly Ile Asp Asp Leu	
290 295 300	
tgc ctg aaa gcg ttt gtc gag gcc aac gtt atg gca gtc cgc cgt tgt	1438
Cys Leu Lys Ala Phe Val Glu Ala Asn Val Met Ala Val Arg Arg Cys	
305 310 315 320	
cag aag gag gat ttg cgc agg ata gct cgc gcc aca ggc gct gtg ctc	1486
Gln Lys Glu Asp Leu Arg Arg Ile Ala Arg Ala Thr Gly Ala Val Leu	
325 330 335	
atc agc tca atg tcg aac ttg gag ggc gat gag acc ttc gag cca agc	1534
Ile Ser Ser Met Ser Asn Leu Glu Gly Asp Glu Thr Phe Glu Pro Ser	
340 345 350	
tat ctc ggc acc tgt agc gag gtc gtg cag gcg aag ttc tcc gac gac	1582
Tyr Leu Gly Thr Cys Ser Glu Val Val Gln Ala Lys Phe Ser Asp Asp	
355 360 365	
gag tgt atc ttg att aag ggc act gcc aag cat tct tct tcc tcc atc	1630
Glu Cys Ile Leu Ile Lys Gly Thr Ala Lys His Ser Ser Ser Ser Ile	
370 375 380	
atc ttg cgt gga cca aac gac tac tct ctg gac gaa atg gag cgc tcg	1678
Ile Leu Arg Gly Pro Asn Asp Tyr Ser Leu Asp Glu Met Glu Arg Ser	
385 390 395 400	
ctg cac gac tcg ctc tgc gtg gtg aag aga acc ctc gag agt ggg cat	1726
Leu His Asp Ser Leu Cys Val Val Lys Arg Thr Leu Glu Ser Gly His	
405 410 415	
gtg gtg gct ggc ggc ggc tgt gtc gag gcg gcc ctc aac ata tat cta	1774
Val Val Ala Gly Gly Gly Cys Val Glu Ala Ala Leu Asn Ile Tyr Leu	
420 425 430	
gac aac ttt gcc acc acc gtc ggc tcg cgc gag cag ctg gca atc gcc	1822
Asp Asn Phe Ala Thr Thr Val Gly Ser Arg Glu Gln Leu Ala Ile Ala	
435 440 445	

gag ttt gcc gcc gcg atg ttg gtc atc cca aag aca ttg gcc gtc aat 1870
 Glu Phe Ala Ala Ala Met Leu Val Ile Pro Lys Thr Leu Ala Val Asn
 450 455 460
 gca gcc aag gac tcc agc gac ctc gtt gcc aag ctg cgt tcg tac cat 1918
 Ala Ala Lys Asp Ser Ser Asp Leu Val Ala Lys Leu Arg Ser Tyr His
 465 470 475 480
 gca gcc agc cag cag gcc aaa ccg gac gca cct caa gaa gcg aag tac 1966
 Ala Ala Ser Gln Gln Ala Lys Pro Asp Ala Pro Gln Glu Ala Lys Tyr
 485 490 495
 cgc aac tac ggg ctc gac ctt atc cgc ggg aag gtc gtc gac gag gtg 2014
 Arg Asn Tyr Gly Leu Asp Leu Ile Arg Gly Lys Val Val Asp Glu Val
 500 505 510
 gga gcc ggc gtg ttg gaa ccc acc atc agc aag aaa aag tct ttg aag 2062
 Gly Ala Gly Val Leu Glu Pro Thr Ile Ser Lys Lys Lys Ser Leu Lys
 515 520 525
 tcc gcc ctc gaa gct tgc att gcc att ctt aga att gac aca atg att 2110
 Ser Ala Leu Glu Ala Cys Ile Ala Ile Leu Arg Ile Asp Thr Met Ile
 530 535 540
 acg gtc gat cca gaa cct cca aag gag gat ccc cac gct cac 2152
 Thr Val Asp Pro Glu Pro Pro Lys Glu Asp Pro His Ala His
 545 550 555
 tgattcggaa tatatatgta catatcgttt ctaccattcg gccttttggc tttgaacttc 2212
 tgaaattcct ccgtagtagc atacgacgca cgaagcggct cttcctgcgt cctcacgggc 2272
 gttcacacag ctctgttata caactgcgca ttctgtatctt accttagcgg atcttctgga 2332
 gaggagatc 2341

<210> 19

<211> 558

<212> PRT

<213> Ashbya gossypii

<220>

<221> misc_feature

<223> Oligo 161

<400> 19

Met Ala Gln Val Phe Asn Asn Gln Arg Ser Asp Thr Leu Phe Leu Gly
 1 5 10 15

Gly Glu Lys Val Ser Gly Asp Asp Ile Arg Asn Gln Asn Val Leu Ala

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

Arg Lys Arg Glu Val Gly Ile Val Leu Glu Arg Val Lys Lys Ile Ile
 275 280 285

Asp Ala Gly Ala Gln Val Val Leu Thr Thr Lys Gly Ile Asp Asp Leu
 290 295 300

Cys Leu Lys Ala Phe Val Glu Ala Asn Val Met Ala Val Arg Arg Cys
 305 310 315 320

Gln Lys Glu Asp Leu Arg Arg Ile Ala Arg Ala Thr Gly Ala Val Leu
 325 330 335

Ile Ser Ser Met Ser Asn Leu Glu Gly Asp Glu Thr Phe Glu Pro Ser
 340 345 350

Tyr Leu Gly Thr Cys Ser Glu Val Val Gln Ala Lys Phe Ser Asp Asp
 355 360 365

Glu Cys Ile Leu Ile Lys Gly Thr Ala Lys His Ser Ser Ser Ser Ile
 370 375 380

Ile Leu Arg Gly Pro Asn Asp Tyr Ser Leu Asp Glu Met Glu Arg Ser
 385 390 395 400

Leu His Asp Ser Leu Cys Val Val Lys Arg Thr Leu Glu Ser Gly His
 405 410 415

Val Val Ala Gly Gly Gly Cys Val Glu Ala Ala Leu Asn Ile Tyr Leu
 420 425 430

Asp Asn Phe Ala Thr Thr Val Gly Ser Arg Glu Gln Leu Ala Ile Ala
 435 440 445

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

Ala Ala Lys Asp Ser Ser Asp Leu Val Ala Lys Leu Arg Ser Tyr His
 465 470 475 480

Ala Ala Ser Gln Gln Ala Lys Pro Asp Ala Pro Gln Glu Ala Lys Tyr
 485 490 495

Arg Asn Tyr Gly Leu Asp Leu Ile Arg Gly Lys Val Val Asp Glu Val
 500 505 510

Gly Ala Gly Val Leu Glu Pro Thr Ile Ser Lys Lys Lys Ser Leu Lys

515

520

525

Ser Ala Leu Glu Ala Cys Ile Ala Ile Leu Arg Ile Asp Thr Met Ile
530 535 540

Thr Val Asp Pro Glu Pro Pro Lys Glu Asp Pro His Ala His
545 550 555