



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 693 34 197 T2** 2008.12.11

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 967 279 B1**

(21) Deutsches Aktenzeichen: **693 34 197.1**

(96) Europäisches Aktenzeichen: **99 202 698.9**

(96) Europäischer Anmeldetag: **02.03.1993**

(97) Erstveröffentlichung durch das EPA: **29.12.1999**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **02.01.2008**

(47) Veröffentlichungstag im Patentblatt: **11.12.2008**

(51) Int Cl.⁸: **C12N 15/31** (2006.01)

A61K 39/106 (2006.01)

C12Q 1/68 (2006.01)

G01N 33/569 (2006.01)

(30) Unionspriorität:

FI920052 02.03.1992 IT
PCT/EP93/00158 25.01.1993 WO

(84) Benannte Vertragsstaaten:

AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE

(73) Patentinhaber:

Novartis Vaccines and Diagnostics S.r.l., Siena, IT

(72) Erfinder:

Covacci, Antonello, 53100 Siena, IT; Bugnoli, Massimo, 53100 Siena, IT; Telford, John, 53100 Siena, IT; Macchia, Giovanni, 53100 Siena, IT; Rappuoli, Rino, 53100 Siena, IT

(74) Vertreter:

Vossius & Partner, 81675 München

(54) Bezeichnung: **Helicobacter pylori Cytotoxin verwendbar in Impfstoffe und Diagnostik**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung**HINTERGRUND DER ERFINDUNG****1. Bereich der Offenbarung**

[0001] Die vorliegende Erfindung betrifft das *Helicobacter pylori*-Protein, das Gen, das dieses Protein exprimiert, und die Verwendung dieses Proteins für diagnostische Anwendung und Impfanwendungen.

2. Kurze Beschreibung des Fachgebietes

[0002] *Helicobacter pylori* ist ein spiralförmiges, mikroaerophiles, Gram-negatives Bakterium, das zum ersten Mal 1982 aus Magenbiopsien von Patienten mit chronischer Gastritis isoliert wurde, Warren et al., *Lancet* i (1983), 1273–75. Ursprünglich *Campylobacter pylori* genannt, wurde es als Teil einer separaten Gattung mit Namen *Helicobacter* erkannt, Goodwin et al., *Int. J. Syst. Bacteriol.* 39 (1989), 397–405. Das Bakterium besiedelt die menschliche Magenschleimhaut, und die Infektion kann über Jahrzehnte andauern. Während der letzten Jahre wurde die Anwesenheit des Bakteriums mit chronischer Typ-B-Gastritis in Verbindung gebracht, eine Erkrankung, die in den meisten infizierten Personen asymptomatisch bleiben kann, jedoch das Risiko eines Magengeschwürs und eines Magen-Adenokarzinoms beträchtlich erhöht. Die jüngsten Studien deuten stark darauf hin, dass eine Infektion mit *H. pylori* entweder eine Ursache oder ein Cofactor für Typ B-Gastritis, Magengeschwüre und Magentumore sein könnte, siehe z. B. Blaser, *Gastroenterology* 93 (1987), 371–83; Dooley et al., *New Engl. J. Med.* 321 (1989), 1562–66; Parsonnet et al., *New Engl. J. Med.* 325 (1991), 1127–31. *H. pylori* wird wahrscheinlich auf oralem Weg übertragen, Thomas et al., *Lancet* i 340 (1992), 1194, und das Risiko einer Infektion steigt mit dem Alter, Graham et al., *Gastroenterology* 100 (1991), 1495–1501, und wird durch Überbevölkerung vergrößert, Drumm et al., *New Engl. J. Med.* 4322 (1990), 359–63; Blaser, *Clin. Infect. Dis.* 15 (1992), 386–93. In den Industrieländern steigt die Anwesenheit von Antikörpern gegen *H. pylori*-Antigene von weniger als 20% auf über 50% in Menschen im Alter von 30 bzw. 60 Jahren an, Jones et al., *Med. Microbio.* 22 (1986), 57–62; Morris et al., *N. Z. Med. J.* 99 (1986), 657–59, während in den Entwicklungsländern über 80% der Bevölkerung bereits im Alter von 20 Jahren infiziert sind, Graham et al., *Digestive Diseases and Sciences* 36 (1991), 1084–88.

[0003] Die Beschaffenheit und die Rolle der Virulenzfaktoren von *H. pylori* werden noch schlecht verstanden. Die Faktoren, die bislang identifiziert wurden, schließen die Flagellen ein, die möglicherweise für die Beweglichkeit in der Mucus-Schicht notwendig sind, siehe z. B. Leying et al., *Mol. Microbiol.* 6 (1992), 2863–74; die Urease, die für die Neutralisie-

rung der sauren Umgebung des Magens notwendig ist und die anfängliche Besiedelung erlaubt, siehe z. B. Cussac et al., *J. Bacteriol.* 174 (1992), 2466–73; Perez-Perez et al., *J. Infect. Immun.* 60 (1992), 3658–3663; Austin et al., *J. Bacteriol.* 174 (1992), 7470–73; PCT-Veröffentlichung Nr. WO 90/04030; und ein hochmolekulares cytotoxisches Protein, das von Monomeren mit einem angegebenen Molekulargewicht von 87 kDa gebildet wird, das die Bildung von Vakuolen in eukaryotischen Epithelzellen verursacht und von *H. pylori*-Stämmen produziert wird, die mit Erkrankungen im Zusammenhang stehen, siehe z. B. Cover et al., *J. Bio. Chem.* 267 (1992), 10570–75 (Hinweis auf ein "vakuolisierendes Toxin" mit einer aufgeführten N-terminalen Sequenz von 23 Aminosäuren); Cover et al., *J. Clin. Invest.* 90 (1992), 913–18; Leunk, *Rev. Infect. Dis.* 13 (1991), 5686–89.

[0004] Für Beispiele diagnostischer Tests auf der Grundlage von *H. pylori*-Lysaten oder teilweise gereinigten Antigenen siehe Evans et al., *Gastroenterology* 96 (1989), 1004–08; U.S. 4,882,271; PCT-Veröffentlichung Nr. WO 89/08843 (alle betreffen Zusammensetzungen und Tests, die die gleichen hochmolekularen Antigene (300–700 kDa) von der Oberfläche der äußeren Membran mit Urease-Aktivität beinhalten); EPO-Veröffentlichung Nr. 329 570 (betrifft ein Mittel aus Antigenen zum Nachweis von *H. pylori*-Antikörpern mit Fragmenten von mindestens einem Fragment aus der Gruppe mit 63, 57, 45, und 31 kDa).

[0005] Der Prozentsatz von Menschen, die mit *H. pylori* infiziert sind, entweder in einer symptomatischen oder asymptomatischen Form, ist sowohl in den Entwicklungsländern als auch in den Industrieländern sehr hoch, und die Kosten für Krankenhausaufenthalt und Therapie machen die Entwicklung von *H. pylori*-Impfstoffen und weiteren diagnostischen Tests für diese Erkrankung wünschenswert.

ZUSAMMENFASSUNG DER ERFINDUNG

[0006] Die vorliegende Erfindung beschreibt die Nucleotid- und Aminosäuresequenz für ein *H. pylori*-Hauptprotein, das Cytotoxin. Die vollständige Aminosäuresequenz für dieses Protein ist nicht bekannt, und das Gen wurde nicht identifiziert. Die vorliegende Erfindung betrifft nicht nur dieses gereinigte Protein und sein Gen, sondern auch damit zusammenhängendes rekombinantes Material, wie Vektoren und Wirtszellen. Das Verständnis der Beschaffenheit und der Rolle dieses Proteins auf molekularer Ebene und die Verfügbarkeit einer rekombinanten Herstellung hat bedeutende Folgen für die Entwicklung neuer Diagnostika für *H. pylori* und für den Entwurf von Impfstoffen, die die Infektion mit *H. pylori* verhindern und die Erkrankung behandeln könnten.

[0007] Dieses Protein kann daher sowohl in Impfun-

gen als auch in diagnostischen Anwendungen verwendet werden. Die vorliegende Erfindung schließt Methoden zur Behandlung und zur Diagnose dieser mit *H. pylori* assoziierten Erkrankungen ein. Da *H. pylori* mit Typ-B-Gastritis, Magengeschwüren, und dem Magen-Adenokarzinom in Zusammenhang gebracht wurde, ist zu hoffen, dass die vorliegende Erfindung zu einer frühen Erkennung und einer Linderung dieser Erkrankungszustände beitragen wird. Gegenwärtig beruht die Diagnose hauptsächlich auf Endoskopie und histologischer Anfärbung von Biopsien; existierende Immuntests basieren auf *H. pylori*-Lysaten oder teilweise gereinigten Antigenen. Angesichts der in solchen Ansätzen gefundenen Heterogenität ist die Korrelation mit den Erkrankungszuständen noch nicht gut etabliert. Das Potential für rekombinante Immuntests auf Basis von Antigenen wie auch Nucleinsäure-Tests zur Erkennung der Erkrankung ist daher groß. Gegenwärtig gibt es im Handel keine Impfung gegen eine *H. pylori*-Infektion oder zur Behandlung. Ein rekombinanter Impfstoff ist daher ein Gegenstand der vorliegenden Erfindung.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0008] **Fig. 1** zeigt die Nucleotidsequenz für das Cytotoxin (CT)-Protein.

[0009] **Fig. 2** zeigt die Aminosäuresequenz für das Cytotoxin (CT)-Protein.

GENAUE BESCHREIBUNG DER ERFINDUNG

A. Allgemeine Methoden

[0010] Die Ausführung der vorliegenden Erfindung verwendet, wenn nicht anders angegeben, konventionelle Techniken aus Molekularbiologie, Mikrobiologie, rekombinanter DNA, und Immunologie, die Stand der Technik des Fachgebietes sind. Diese Techniken werden ausführlich in der Literatur abgehandelt. Siehe z. B. Sambrook et al. (1989), *MOLECULAR CLONING; A LABORATORY MANUAL* (ZWEITE AUFLAGE); Glover, D. N. (Hrsg.) (1985), *DNA CLONING, BÄNDE I UND II*; Gait, M. J. (Hrsg.) (1984), *OLIGONUCLEOTIDE SYNTHESIS*; Hames B. D. & Higgins S. J. (Hrsg.) (1984), *NUCLEIC ACID HYBRIDIZATION*; Hames B. D. & Higgins S. J. (Hrsg.) (1984), *TRANSCRIPTION AND TRANSLATION*; Freshney, R. I. (Hrsg.) (1986), *ANIMAL CELL CULTURE; IMMOBILIZED CELLS AND ENZYMES* (1986), IRL Press; Perbal, B. (1984), *A PRACTICAL GUIDE TO MOLECULAR CLONING*; die Serie *METHODS IN ENZYMOLOGY*, Academic Press, Inc.; Miller, J. H. and Calos, M. P. (Hrsg.) (1987), *GENE TRANSFER VECTORS FOR MAMMALIAN CELLS*, Cold Spring Harbor Laboratory, Wu and Grossman (Hrsg.) bzw. Wu (Hrsg.), *Methods in Enzymology*, Vol. 154 bzw. Vol. 155, Mayer and Walker (Hrsg.) (1987), *IMMUNOCHEMICAL METHODS IN CELL*

AND MOLECULAR BIOLOGY, London: Academic Press, Scopes (1987), *PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE* (Zweite Auflage), N. Y: Springer Verlag, und Weir, D. M. and Blackwell, C. C. (Hrsg.) (1986), *HANDBOOK OF EXPERIMENTAL IMMUNOLOGY*, BÄNDE I-IV.

[0011] In dieser Patentbeschreibung werden die Standardabkürzungen für Nucleotide und Aminosäuren verwendet.

B. Definitionen

[0012] "Cytotoxin" oder "Toxin" von *H. pylori* bezeichnet das Protein, und Fragmente davon, dessen Nucleotidsequenz und Aminosäuresequenzen in **Fig. 1** bzw. **Fig. 2** gezeigt werden, und ihre Derivate, und dessen Molekulargewicht ungefähr 140 kDa beträgt. Dieses Protein dient als Vorläufer eines Proteins mit einem ungefähren Molekulargewicht von 100 kDa und mit cytotoxischer Aktivität. Das Cytotoxin verursacht Vakuolisierung und Tod einer Anzahl eukaryotischer Zelltypen und wurde aus *H. pylori*-Kulturüberständen aufgereinigt. Darüberhinaus ist das Cytotoxin proteinös und hat ein scheinbares, durch Gelfiltration bestimmtes Molekulargewicht von ungefähr 950–972 kDa. Denaturierende Gelelektrophorese von gereinigtem Material hatte früher ergeben, dass die Hauptkomponente des 950–972 kDa Moleküls angeblich ein Polypeptid mit einem scheinbaren Molekulargewicht von 87 kDa war, Cover et al., *J. Biol. Chem.* 267 (1992), 10570–75. Es wird jedoch hier vorgeschlagen, dass das ursprünglich beschriebene 87 kDa Protein entweder aus einer weiteren Prozessierung des 100 kDa Proteins oder aus dem proteolytischen Abbau eines größeren Proteins während der Aufreinigung hervorgeht.

[0013] Beispiele von Proteinen, die in der vorliegenden Erfindung verwendet werden können, schließen Polypeptide mit geringfügigen Aminosäurevariationen der natürlichen Aminosäuresequenz des Proteins ein; insbesondere konservative Aminosäureaustausche werden betrachtet. Konservative Austausche sind solche, die innerhalb einer in ihren Seitenketten verwandten Familie ablaufen. Genetisch codierte Aminosäuren sind im Allgemeinen in vier Familien unterteilt: (1) sauer = Aspartat, Glutamat; (2) basisch = Lysin, Arginin, Histidin; (3) unpolar = Alanin, Valin, Leucin, Isoleucin, Prolin, Phenylalanin, Methionin, Tryptophan; und (4) ungeladen polar = Glycin, Asparagin, Glutamin, Cystein, Serin, Threonin, Tyrosin. Phenylalanin, Tryptophan, und Tyrosin werden manchmal gemeinsam als aromatische Aminosäuren klassifiziert. Zum Beispiel ist es ziemlich vorhersagbar, dass ein isolierter Austausch eines Leucins gegen ein Isoleucin oder Valin, eines Aspartats gegen ein Glutamat, eines Threonins gegen ein Serin, oder ein ähnlicher konservativer Austausch einer Aminosäure mit einer strukturell verwandten Aminosäure

keinen wesentlichen Einfluss auf die biologische Aktivität haben wird. Polypeptidmoleküle mit einer im Wesentlichen gleichen Aminosäuresequenz wie das Protein, aber mit geringfügigen Aminosäureaustauschen, die die funktionellen Aspekte nicht grundlegend beeinflussen, liegen innerhalb der Definition des Proteins.

[0014] Ein wesentlicher Vorteil der Produktion des Proteins durch rekombinante DNA-Techniken gegenüber der Isolierung und Reinigung eines Proteins aus natürlichen Quellen besteht darin, dass gleichwertige Proteinmengen aus weniger Ausgangsmaterial produziert werden können, als für eine Isolierung des Proteins aus einer natürlichen Quelle erforderlich wäre. Die Produktion eines Proteins durch rekombinante Techniken erlaubt auch die Isolierung des Proteins in Abwesenheit einiger Moleküle, die normalerweise in Zellen vorhanden sind. In der Tat können leicht Proteinpräparationen hergestellt werden, die gänzlich frei von jeder Spur humaner Protein-Kontaminanten sind, weil das einzige Humanprotein, das von der rekombinanten nicht-humanen Wirtszelle hergestellt wird, das betreffende rekombinante Protein ist. Mögliche virale Wirkstoffe aus natürlichen Quellen und für den Menschen pathogene virale Bestandteile werden dadurch ebenso vermieden.

[0015] Der Begriff "rekombinantes Polynucleotid", wie hier verwendet, meint ein Polynucleotid genomischen, cDNA-, teilsynthetischen, oder synthetischen Ursprungs, das aufgrund seines Ursprungs oder Manipulation: (1) nicht mit dem vollständigen oder einem Teil des Polynucleotides verbunden ist, mit dem es in der Natur verbunden ist, (2) mit einem anderen Polynucleotid verbunden ist als demjenigen, mit dem es in der Natur verbunden ist, oder (3) in der Natur nicht vorkommt. Dieser Begriff schließt daher auch den Umstand ein, in dem das *H. pylori*-Bakteriengenom genetisch verändert wird (z. B. durch Mutagenese), um ein oder mehrere veränderte Polypeptide herzustellen.

[0016] Der Begriff "Polynucleotid", wie hier verwendet, bezeichnet die polymere Form eines Nucleotids jeglicher Länge, bevorzugt Desoxyribonucleotide, und wird hier austauschbar mit den Begriffen "Oligonucleotid" und "Oligomer" verwendet. Der Begriff bezieht sich nur auf die Primärstruktur des Moleküls. Der Begriff schließt daher doppel- und einzelsträngige DNA, sowie Antisense-Polynucleotide ein. Er schließt auch bekannte Arten der Modifikation ein, zum Beispiel, die Anwesenheit von Markierungen, die auf dem Fachgebiet bekannt sind, Methylierung, Kopfgruppen (End-"Caps"), Ersatz von einem oder mehreren der natürlich vorkommenden Nucleotide durch ein Analogon, Internucleotid-Modifikationen wie, zum Beispiel, Austausch gegen bestimmte Typen ungeladener Verknüpfungen (z. B. Methylphosphonate, Phosphotriester, Phosphoramidate, Carba-

mate, usw.) oder geladener Verknüpfungen (z. B. Phosphorothioate, Phosphorodithioate, usw.), Einführung hängender Einheiten, wie, zum Beispiel, Proteine (eingeschlossen Nucleasen, Toxine, Antikörper, Signalpeptide, Poly-L-Lysin, usw.), Intercalatoren (z. B. Acridin, Psoralen, usw.), Chelatoren (z. B. Metalle, radioaktive Spezies, Bor, oxidative Einheiten, usw.), alkylierende Verbindungen (z. B. alpha-anomerische Nucleinsäuren, usw.).

[0017] Mit "genomisch" ist eine Sammlung oder eine Genbank von DNA-Molekülen gemeint, die sich aus Restriktionsfragmenten ableiten, die in Vektoren cloniert wurden. Dies kann das gesamte oder einen Teil des genetischen Materials eines Organismus einschließen.

[0018] Mit "cDNA" ist eine komplementäre mRNA-Sequenz gemeint, die mit einem komplementären mRNA-Strang hybridisiert.

[0019] Der Begriff "Oligomer", so wie hier verwendet, bezieht sich sowohl auf Primer als auch auf Sonden und wird hier austauschbar mit dem Begriff "Polynucleotid" verwendet. Der Begriff Oligomer bedeutet nicht zugleich die Größe des Moleküls. Typischerweise sind Oligomere jedoch nicht länger als 1000 Nucleotide, typischer nicht länger als 500 Nucleotide, noch typischer nicht länger als 250 Nucleotide; sie können nicht länger als 100 Nucleotide, und sie können nicht länger als 75 Nucleotide sein, und sie können auch nicht länger als 50 Nucleotide sein.

[0020] Der Begriff "Primer", so wie hier verwendet, bezeichnet ein Oligomer, das als Startpunkt für die Synthese eines Polynucleotidstrangs bei Anwendung geeigneter Reaktionsbedingungen fungieren kann. Der Primer ist vollständig oder im Wesentlichen komplementär zu einem Bereich des Polynucleotidstranges, der kopiert werden soll. Der Primer lagert sich daher unter Bedingungen, die eine Hybridisierung begünstigen, an den komplementären Bereich des Analytstrangs an. Nach Zugabe geeigneter Reaktionspartner (z. B. eine Polymerase, Nucleotidtriphosphate, und ähnliche), wird der Primer durch den polymerisierenden Wirkstoff verlängert, um eine Kopie des Analytstrangs zu bilden. Der Primer kann einzelsträngig oder in einer anderen Ausführungsform auch teilweise oder vollständig doppelsträngig sein.

[0021] Die Begriffe "Analytpolynucleotid" und "Analytstrang" beziehen sich auf ein einzel- oder doppelsträngiges Nucleinsäuremolekül, von dem angenommen wird, dass es eine Zielsequenz enthält, und das in einer biologischen Probe vorhanden sein kann.

[0022] Wie hier verwendet, bezieht sich der Begriff "Sonde" auf eine aus einem Polynucleotid bestehende Struktur, die aufgrund der Komplementarität wenigstens einer Sequenz in der Sonde mit einer Se-

quenz im Bereich des Zieles eine Hybridstruktur mit einer Zielsequenz bildet. Die Polynucleotidbereiche der Sonden können sich aus DNA, und/oder RNA, und/oder synthetischen Nucleotidanaloga zusammensetzen. Eingeschlossen in die Sonden sind "Fangsonden" und "Markierungssonden".

[0023] Der Begriff "Zielbereich", so wie hier verwendet, bezieht sich auf einen Bereich der Nucleinsäure, der amplifiziert und/oder detektiert werden soll. Der Begriff "Zielsequenz" bezieht sich auf eine Sequenz, mit der eine Sonde oder ein Primer ein stabiles Hybrid unter den erwünschten Bedingungen bildet.

[0024] Der Begriff "Fangsonde", so wie hier verwendet, bezieht sich auf eine Polynucleotidsonde, die ein mit einem Bindungspartner gekoppeltes einzelsträngiges Polynucleotid enthält. Das einzelsträngige Polynucleotid enthält eine Ziel-Polynucleotidsequenz, die komplementär zu einer Zielsequenz in einem Zielbereich ist, die in dem Analytpolynucleotid nachgewiesen werden soll. Dieser komplementäre Bereich ist ausreichend lang und komplementär zur Zielsequenz, um einen stabilen Doppelstrang zu ermöglichen, der ausreichend ist, um das Analytpolynucleotid an einer festen Oberfläche (über die Bindungspartner) zu immobilisieren. Der Bindungspartner ist spezifisch für einen zweiten Bindungspartner; der zweite Bindungspartner kann an die Oberfläche eines festen Trägers gebunden sein, oder kann indirekt über andere Strukturen oder Bindungspartner mit einem festen Träger verbunden sein.

[0025] Der Begriff "zielende Polynucleotidsequenz", so wie hier verwendet, bezieht sich auf eine Polynucleotidsequenz, die Nucleotide enthält, die komplementär zu einer Ziel-Nucleotidsequenz sind; die Sequenz ist ausreichend lang und komplementär zur Zielsequenz, um einen Doppelstrang zu bilden, der eine ausreichende Stabilität für den vorgesehenen Zweck hat.

[0026] Der Begriff "Bindungspartner", so wie hier verwendet, bezieht sich auf ein Molekül, das in der Lage ist, ein Ligandenmolekül mit hoher Spezifität zu binden, wie zum Beispiel ein Antigen und ein dafür spezifischer Antikörper. Im Allgemeinen müssen die spezifischen Bindungspartner mit ausreichender Affinität binden, um den Doppelstrang aus Analytkopie/Komplementärstrang (im Falle der Fangsonden) unter den Bedingungen der Isolierung zu immobilisieren. Spezifische Bindungspartner sind auf dem Fachgebiet bekannt, und schließen, zum Beispiel, Biotin und Avidin oder Streptavidin, IgG und Protein A, die zahlreichen Rezeptor-Liganden-Paare, und komplementäre Polynucleotidstränge ein. Im Falle von komplementären Polynucleotid-Bindungspartnern sind die Partner normalerweise mindestens 15 Basen lang, und können mindestens 40 Basen lang sein; zusätzlich haben sie einen G- und C-Gehalt von min-

destens ungefähr 40% bis hin zu ungefähr 60%. Das Polynucleotid kann aus DNA, RNA oder synthetischen Nucleotidanaloga bestehen.

[0027] Der Begriff "gekoppelt", wie hier verwendet, bezieht sich auf die Verknüpfung durch kovalente Bindungen oder durch starke nicht-kovalente Wechselwirkungen (z. B. hydrophobe Wechselwirkungen, Wasserstoffbrückenbindungen, usw.). Kovalente Bindungen können zum Beispiel Ester-, Ether-, Phosphorester-, Amid-, Peptid-, Imid-, Kohlenstoff-Schwefel-Bindungen, Kohlenstoff-Phosphor-Bindungen, und ähnliche sein.

[0028] Der Begriff "Träger" bezieht sich auf jede feste oder halb feste Oberfläche, an der ein gewünschter Bindungspartner verankert sein kann. Geeignete Träger schließen Glas, Plastik, Metall, Polymer-Gele und ähnliche ein, und können die Form von Kugeln, Vertiefungen, Messstäben, Membranen, und ähnliche haben.

[0029] Der Begriff "Markierung", wie hier verwendet, bezieht sich auf jedes Atom oder jede Einheit, das oder die zur Erzeugung eines nachweisbaren (vorzugsweise quantifizierbaren) Signals verwendet und mit einem Polynucleotid oder Polypeptid verknüpft werden kann.

[0030] Wie hier verwendet, bezieht sich der Begriff "Markierungssonde" auf eine Polynucleotidsonde, die eine zielsuchende Polynucleotidsequenz enthält, die zu einer im Analytpolynucleotid nachzuweisen Zielsequenz komplementär ist. Dieser komplementäre Bereich ist ausreichend lang und komplementär zur Zielsequenz, um einen Doppelstrang aus der "Markierungssonde" und der "Zielsequenz" zu ermöglichen. Die Markierungssonde ist entweder direkt mit einer Markierung, oder indirekt über eine Gruppe von Ligandenmolekülen mit hoher Spezifität für einander, eingeschlossen Multimere, gekoppelt.

[0031] Der Begriff "Multimer", wie hier verwendet, bezieht sich auf lineare oder verzweigte Polymere der gleichen, sich wiederholenden einzelsträngigen Polynucleotideinheit oder verschiedener einzelsträngiger Polynucleotideinheiten. Mindestens eine der Einheiten hat eine Sequenz, eine Länge und eine Zusammensetzung, die es ihr erlaubt, spezifisch an eine erste einzelsträngige Nucleotidsequenz von Interesse zu hybridisieren, typischerweise an einen Analyt oder an eine an einen Analyt gebundene Polynucleotidsonde (z. B. eine Markierungssonde). Um eine solche Spezifität und Stabilität zu ermöglichen, ist diese Einheit mindestens ungefähr 15 Nucleotide lang, typischerweise nicht mehr als 50 Nucleotide lang, und bevorzugt ungefähr 30 Nucleotide lang; darüberhinaus beträgt der G- und C-Gehalt normalerweise mindestens ungefähr 40% und höchstens ungefähr 60%. Zusätzlich zu einer solchen Einheit

(solchen Einheiten) schließt das Multimer eine Vielzahl von Einheiten ein, die in der Lage sind, spezifisch und stabil an ein zweites einzelsträngiges Nucleotid von Interesse, typischerweise ein markiertes Polynucleotid oder ein anderes Multimer, zu hybridisieren. Diese Einheiten haben im Allgemeinen etwa die gleiche Größe und Zusammensetzung wie die oben angeführten Multimere. Wenn ein Multimer entworfen wird, um mit einem anderen Multimer hybridisiert zu werden, sind die erste und die zweite Oligonucleotideinheit heterogen (verschieden), und hybridisieren unter den Bedingungen des ausgewählten Assays nicht miteinander. Daher können Multimere Markierungs sonden sein, oder Liganden, die die Markierung an die Sonde koppeln.

[0032] Ein "Replikon" ist jedes genetische Element, z. B. ein Plasmid, ein Chromosom, ein Virus, ein Cosmid, usw., das sich wie eine eigenständig replizierende Polynucleotideinheit innerhalb einer Zelle verhält; d. h., befähigt zur Replikation unter seiner eigenen Kontrolle. Dies kann selektierbare Marker einschließen.

[0033] "PCR" bezieht sich auf die Technik der Polymerase-Kettenreaktion wie beschrieben in Saiki et al., Nature 324 (1986), 163; und Scharf et al., Science 233 (1986), 1076–1078; und US 4,683,195; und US 4,683,202.

[0034] Wie hier verwendet, ist x "heterolog" in Bezug auf y, wenn x nicht natürlicherweise mit y in der identischen Weise verbunden ist; d. h., x ist in der Natur nicht mit y verbunden oder x ist nicht in der gleichen Weise mit y wie in der Natur verbunden.

[0035] "Homologie" bezieht sich auf den Grad der Ähnlichkeit zwischen x und y. Die Übereinstimmung zwischen der Sequenz einer Form mit einer anderen kann durch im Fachgebiet bekannte Techniken bestimmt werden. Zum Beispiel kann sie durch direkten Vergleich der Sequenzinformation des Polynucleotids bestimmt werden. Alternativ kann die Homologie durch Hybridisierung der Polynucleotide unter Bedingungen bestimmt werden, bei denen sich stabile Doppelstränge zwischen homologen Bereichen ausbilden (zum Beispiel solche, die vor einem S1-Verdau verwendet würden), gefolgt von einer Spaltung mit einer für Einzelstränge spezifischen Nuclease (mit für Einzelstränge spezifischen Nucleasen), gefolgt von einer Größenbestimmung der gespaltenen Fragmente.

[0036] Ein "Vektor" ist ein Replikon, mit dem ein anderes Polynucleotidsegment verbunden ist, so dass die Replikation und/oder Expression des verbundenen Segmentes bewerkstelligt wird.

[0037] "Kontrollsequenz" bezieht sich auf Polynucleotidsequenzen, die erforderlich sind, um eine Ex-

pression der codierenden Sequenzen zu bewirken, mit denen sie ligiert sind. Die Beschaffenheit solcher Kontrollsequenzen differiert in Abhängigkeit vom Wirtsorganismus; in Prokaryoten schließen solche Kontrollsequenzen meistens einen Promotor, eine Ribosomen-Bindungsstelle, und eine Transkriptions-Terminationssequenz ein; in Eukaryoten schließen solche Kontrollsequenzen meistens Promotoren und eine Transkriptions-Terminationssequenz ein. Der Begriff "Kontrollsequenz" soll mindestens alle Komponenten einschließen, deren Anwesenheit für eine Expression erforderlich ist, und kann auch zusätzliche Komponenten einschließen, deren Anwesenheit vorteilhaft ist, zum Beispiel Signalsequenzen und Fusionspartnersequenzen.

[0038] "Funktionell verbunden" bezieht sich auf eine räumliche Nähe, in der sich die so beschriebenen Komponenten in einer Beziehung befinden, die es ihnen erlaubt, in der geplanten Weise zu funktionieren. Eine mit einer codierenden Sequenz "funktionell verbundene" Kontrollsequenz ist auf solche Art ligiert, dass die Expression der codierenden Sequenz unter Bedingungen erreicht wird, die mit der Kontrollsequenz kompatibel sind.

[0039] Ein "offener Leserahmen" (ORF) ist der Bereich einer Polynucleotidsequenz, der ein Polypeptid codiert; dieser Bereich kann einen Teil der codierenden Sequenz oder die vollständige codierende Sequenz repräsentieren.

[0040] Eine "codierende Sequenz" ist eine Polynucleotidsequenz, die in ein Polypeptid übersetzt wird, normalerweise über eine mRNA, wenn sie unter die Kontrolle geeigneter regulatorischer Sequenzen gestellt wird. Die Grenzen der codierenden Sequenz werden durch ein Translations-Startcodon am 5'-Ende und ein Translations-Stopcodon am 3'-Ende bestimmt. Eine codierende Sequenz kann cDNA und rekombinante Polynucleotidsequenzen einschließen, ist aber nicht darauf beschränkt.

[0041] Wie hier verwendet, bezieht sich der Begriff "Polypeptid" auf ein Polymer aus Aminosäuren und bezieht sich nicht auf eine bestimmte Länge des Produkts; daher sind Peptide, Oligopeptide und Proteine in die Definition von Polypeptid eingeschlossen. Dieser Begriff bezieht sich auch nicht auf Modifikationen des Polypeptids nach der Expression oder schließt diese aus, zum Beispiel Glykosylierungen, Acetylierungen, Phosphorylierungen und dergleichen. Eingeschlossen in die Definition sind, zum Beispiel, Polypeptide, die ein oder mehrere Analoga einer Aminosäure enthalten (eingeschlossen, zum Beispiel, nicht natürliche Aminosäuren, usw.), Polypeptide mit substituierten Bindungen, wie auch andere im Fachgebiet bekannte Modifikationen, die sowohl natürlich als auch nicht natürlich vorkommen.

[0042] Ein Polypeptid oder eine Aminosäuresequenz "abgeleitet von" einer bestimmten Nucleinsäuresequenz bezieht sich auf ein Polypeptid, das eine Sequenz identisch zu der eines in dieser Sequenz codierten Polypeptids besitzt, oder einen Teil davon, wobei der Teil aus mindestens 3–5 Aminosäuren besteht, und stärker bevorzugt mindestens 8–10 Aminosäuren, und besonders bevorzugt mindestens 11–15 Aminosäuren, oder das mit einem in dieser Sequenz codierten Polypeptid immunologisch identifizierbar ist. Diese Terminologie schließt auch ein durch eine bestimmte Nucleinsäuresequenz exprimiertes Polypeptid ein.

[0043] "Immunogen" bezieht sich auf die Fähigkeit eines Polypeptids, eine humorale und/oder zelluläre Immunantwort hervorzurufen, entweder alleine oder verbunden mit einem Träger, in Gegenwart oder Abwesenheit eines Adjuvans. "Neutralisierung" bezieht sich auf eine Immunantwort, die die Infektiosität eines infektiösen Wirkstoffes entweder teilweise oder vollständig blockiert.

[0044] "Epitop" bezieht sich auf eine antigene Determinante eines Peptids, Polypeptids oder Proteins; ein Epitop kann drei oder mehr Aminosäuren in einer dem Epitop eigenen räumlichen Konformation enthalten. Im Allgemeinen besteht ein Epitop aus mindestens 5 solcher Aminosäuren und, eher üblich, aus mindestens 8–10 solcher Aminosäuren. Methoden zur Bestimmung der räumlichen Konformation von Aminosäuren sind auf dem Fachgebiet bekannt und schließen, zum Beispiel, Röntgenkristallographie und zweidimensionale Kernmagnetresonanz ein. Antikörper, die das gleiche Epitop erkennen, können in einem einfachen Immuntest bestimmt werden, der die Fähigkeit eines Antikörpers zeigt, die Bindung eines anderen Antikörpers an ein Zielprotein zu blockieren.

[0045] "Behandlung", wie hier verwendet, bezieht sich auf Prophylaxe und/oder Therapie (d. h. die Modulation jeglicher Krankheitssymptome). Ein "Individuum" weist auf ein Tier hin, das empfänglich für eine Infektion mit *H. pylori* ist und schließt Primaten, eingeschlossen Menschen, ein, ist aber nicht darauf beschränkt. Ein "Impfstoff" ist ein immunogenes, oder anderweitig befähigtes, einen Schutz, ob teilweise oder vollständig, gegen *H. pylori* hervorrufendes, zur Behandlung eines Individuums verwendbare Mittel.

[0046] Die *H. pylori*-Proteine können für die Herstellung von entweder monoklonalen oder polyclonalen Antikörpern verwendet werden, die spezifisch für die Proteine sind. Die Verfahren zur Herstellung dieser Antikörper sind auf dem Fachgebiet bekannt.

[0047] "Rekombinante Wirtszellen", "Wirtszellen", "Zellen", "Zellkulturen" und ähnliche Begriffe bezeichnen, zum Beispiel, Mikroorganismen, Insektenzellen

und Säugerzellen, die als Empfänger für einen rekombinanten Vektor oder andere übertragene DNA verwendet werden können oder verwendet wurden, und schließen die Nachkommen der ursprünglich transformierten Zelle ein. Es versteht sich, dass die Nachkommen einer einzelnen Mutterzelle aufgrund von natürlichen, zufälligen oder absichtlichen Mutationen nicht notwendigerweise in Morphologie oder der genomischen oder gesamten DNA vollständig identisch zur ursprünglichen Mutterzelle sein müssen. Beispiele für Säuger-Wirtszellen schließen Eierstockzellen Chinesischer Hamster (CHO) und Affen-nierenzellen (COS) ein.

[0048] "Zelllinie", wie hier verwendet, bezieht sich spezifisch auf eine Population von Zellen, die zu beständigem oder verlängertem Wachstum und Teilung in vitro befähigt sind. Häufig sind Zelllinien von einer einzelnen Vorläuferzelle abgeleitete clonale Populationen. Auf dem Fachgebiet ist weiterhin bekannt, dass spontane oder induzierte Veränderungen im Karyotyp während der Lagerung oder des Transfers solcher clonaler Populationen auftreten können. Daher müssen Zellen, die sich von der bezeichneten Zelllinie ableiten, nicht exakt identisch mit den Vorläuferzellen oder -kulturen sein, und die bezeichnete Zelllinie enthält solche Varianten. Der Begriff "Zelllinie" schließt auch immortalisierte Zellen ein. Zelllinien schließen bevorzugt nicht-hybride Zelllinien oder Hybridome aus nur zwei Zellarten ein.

[0049] Wie hier verwendet, schließt der Begriff "Mikroorganismen" prokaryotische und eukaryotische mikrobielle Spezies ein, wie Bakterien und Pilze, wobei letztere Hefen und filamentöse Pilze einschließen.

[0050] Wie hier verwendet, bezieht sich "Transformation" auf die Insertion eines exogenen Polynucleotids in eine Wirtszelle, ungeachtet des für die Insertion verwendeten Verfahrens, zum Beispiel direkte Aufnahme, Transduktion, Übertragung durch F-Plasmide oder Elektroporation. Das exogene Polynucleotid kann als nicht-integrierter Vektor, zum Beispiel als Plasmid erhalten bleiben, oder kann in einer anderen Ausführungsform in das Wirtsgenom integriert werden.

[0051] "Gereinigt" und "isoliert" mit Bezug auf eine Polypeptid- oder Nucleotidsequenz bedeutet, dass das angegebene Molekül praktisch frei von anderen biologischen Makromolekülen der gleichen Art vorliegt. Der Begriff "gereinigt", wie hier verwendet, bedeutet die Anwesenheit von bevorzugt mindestens 75 Gew.-%, stärker bevorzugt mindestens 85 Gew.-%, noch stärker bevorzugt mindestens 95 Gew.-% und besonders bevorzugt mindestens 98 Gew.-% biologischer Makromoleküle der gleichen Art (aber Wasser, Puffer und andere kleine Moleküle, insbesondere Moleküle mit einem Molekulargewicht

von weniger als 1000, können anwesend sind).

C. Nucleinsäure-Tests

[0052] Unter Verwendung des Genoms von *H. pylori* als Grundlage können Polynucleotid-Sonden von ungefähr 8 Nucleotiden oder mehr hergestellt werden, die mit dem (den) positiven Strang (Strängen) der RNA oder ihrem Komplement, wie auch mit cDNAs hybridisieren. Diese Polynucleotide dienen als Sonden zum Nachweis, zur Isolierung und/oder Markierung von Polynucleotiden, die Nucleotidsequenzen enthalten, und/oder als Primer für die Transkription und/oder Replikation der Zielsequenzen. Jede Sonde enthält eine zielsuchende Polynucleotidsequenz, die Nucleotide enthält, die zur Ziel-Nucleotidsequenz komplementär sind; die Sequenz ist ausreichend lang und komplementär zur Sequenz, um einen für die vorgesehene Anwendung genügend stabilen Doppelstrang zu bilden. Wenn zum Beispiel der Zweck in der Isolierung eines Analyten, der eine Zielsequenz enthält, durch Immobilisierung besteht, enthält die Sonde einen Polynucleotidbereich, der ausreichend lang und komplementär zur Zielsequenz ist, um eine ausreichende Doppelstrangstabilität zu erlauben, um den Analyten unter den Bedingungen der Isolierung an einer festen Oberfläche zu immobilisieren. Auch wenn die Polynucleotidsonden zum Beispiel als Primer für eine Transkription und/oder Replikation von Zielsequenzen dienen sollen, enthalten die Sonden einen Polynucleotidbereich, der ausreichend lang und komplementär zur Zielsequenz ist, um die Replikation zu ermöglichen. Auch wenn die Polynucleotidsonden zum Beispiel als Markierungs-sonden oder zur Bindung an Multimeren verwendet werden sollen, würde die zielsuchende Nucleotidsequenz ausreichend lang und komplementär sein, um stabile Hybrid-Doppelstränge mit den Markierungs-sonden und/oder Multimeren zu bilden, um den Nachweis des Doppelstrangs zu ermöglichen. Die Sonden können ein Minimum von etwa 4 zusammenhängenden Nucleotiden beinhalten, die komplementär zur Zielsequenz sind; üblicherweise beinhalten die Oligomere ein Minimum von etwa 8 zusammenhängenden Nucleotiden, die komplementär zur Zielsequenz sind, und beinhalten bevorzugt ein Minimum von etwa 14 zusammenhängenden Nucleotiden, die komplementär zur Zielsequenz sind.

[0053] Die Sonden hingegen müssen nicht nur aus der Sequenz bestehen, die komplementär zur Zielsequenz ist. Sie können zusätzliche Nucleotidsequenzen oder andere Einheiten enthalten. Wenn die Sonden zum Beispiel als Primer für die Amplifizierung von Sequenzen über PCR verwendet werden sollen, können sie Sequenzen enthalten, die im Doppelstrang Erkennungsstellen für Restriktionsenzyme bilden, die die Clonierung der amplifizierten Sequenzen erleichtern. Auch wenn zum Beispiel die Sonden als "Fangsonden" in Hybridisierungsansätzen verwendet

werden sollen, sind sie an einen "Bindungspartner", wie oben definiert, gekoppelt. Hergestellt werden die Sonden durch auf dem Fachgebiet bekannte Verfahren, eingeschlossen, zum Beispiel, durch Verfahren, die Ausschneiden, Transkription oder chemische Synthese einschließen.

D. Expressionssysteme

[0054] Sobald die entsprechende codierende Sequenz von *H. pylori* isoliert ist, kann sie in einer Vielzahl verschiedener Expressionssysteme exprimiert werden; zum Beispiel in solchen, die Säugerzellen, Baculoviren, Bakterien und Hefen verwenden.

i. Säuger-Systeme

[0055] Säuger-Expressionssysteme sind auf dem Fachgebiet bekannt. Ein Säuger-Promotor ist jede DNA-Sequenz, die RNA-Polymerase aus einem Säuger binden und stromabwärts (3') die Transkription einer codierenden Sequenz (z. B. eines Struktur-gens) in mRNA initiieren kann. Ein Promotor hat eine Transkriptions-Initiationsregion, die üblicherweise nahe dem 5'-Ende der codierenden Sequenz lokalisiert ist, und eine TATA-Box, die üblicherweise 25–30 Basenpaare (bp) stromaufwärts der Transkriptions-Initiationsstelle liegt. Es wird angenommen, dass die TATA-Box die RNA-Polymerase II lenkt, um die RNA-Synthese an der richtigen Stelle zu beginnen. Ein Säuger-Promotor enthält auch ein stromaufwärts gelegenes Promotorelement, das üblicherweise innerhalb von 100 bis 200 bp stromaufwärts der TATA-Box liegt. Ein stromaufwärts gelegenes Promotor-Element bestimmt die Rate, mit der die Transkription initiiert wird und kann in beiden Orientierungen funktionieren, Sambrook et al. (1989), *Molecular Cloning: A Laboratory Manual* (2. Auflage).

[0056] Gene von Säuger-Viren sind oft hoch exprimiert und haben ein breites Wirtsspektrum; daher liefern Sequenzen, die Gene von Säuger-Viren codieren, besonders nützliche Promotorsequenzen. Beispiele schließen den frühen Promotor von SV40, den LTR-Promotor des Maus-Mamma-Tumor-Virus, den späten Haupt-Promotor von Adenovirus (Ad MLP) und den Herpes simplex-Viruspromotor ein. Zusätzlich liefern von nicht-viralen Genen abgeleitete Sequenzen, wie das murine Metallothionein-Gen, nützliche Promotorsequenzen. Die Expression kann entweder konstitutiv oder reguliert (induzierbar) sein, kann in Abhängigkeit vom Promotor in Hormon-responsiven Zellen mit Glucocorticoid induziert werden.

[0057] Die Anwesenheit eines Enhancerelements (Enhancer), zusammen mit den oben beschriebenen Promotorelementen, erhöht üblicherweise das Expressionsniveau. Ein Enhancer ist eine regulatorische DNA-Sequenz, die die Transkription bis zu

1000fach stimulieren kann, wenn sie mit homologen oder heterologen Promotoren verbunden ist, wobei die Synthese am normalen RNA-Startpunkt beginnt. Enhancer sind auch aktiv, wenn sie stromaufwärts oder stromabwärts der Transkriptions-Initiationsstelle lokalisiert sind, entweder in normaler oder umgekehrter Orientierung, oder in einer Entfernung von mehr als 1000 Nucleotiden vom Promotor entfernt, Maniatis et al., *Science* 236 (1989), 1237; Alberts et al. (1989), *Molecular Biology of the Cell* (2. Auflage). Von Viren abgeleitete Enhancerelemente können besonders nützlich sein, da sie üblicherweise ein breiteres Wirtsspektrum haben. Beispiele schließen den Enhancer für das frühe Gen von SV40, Dijkema et al., *EMBO J.* 4 (1985), 761, und den vom Long Terminal Repeat (LTR) des Rous-Sarkoma-Virus, Gorman et al., *Proc. Natl. Acad. Sci. USA* 79 (1982), 6777, und vom humanen Cytomegalievirus, Boshart et al., *Cell* 41 (1985), 5221, abgeleiteten Enhancer/Promotor ein. Zusätzlich sind manche Enhancer regulierbar und werden erst in der Anwesenheit eines Induktors, wie eines Hormons oder eines Metallions, aktiv, Sassone-Corsi et al., *Trends Genet.* 2 (1986), 215; Maniatis et al., *Science* 236 (1987), 1237.

[0058] Ein DNA-Molekül kann intrazellulär in Säugerzellen exprimiert werden. Eine Promotorsequenz kann direkt mit dem DNA-Molekül verbunden sein, wobei dann die erste Aminosäure am N-Terminus des rekombinanten Proteins immer ein Methionin ist, das durch das ATG-Startcodon codiert wird. Falls erforderlich, kann der N-Terminus vom Protein durch in vitro-Inkubation mit Bromcyan abgespalten werden.

[0059] In einer anderen Alternative können Fremdproteine auch durch Erzeugung chimärer DNA-Moleküle, die ein Fusionsprotein codieren, das ein Leadersequenz-Fragment enthält, das für die Sekretion des Fremdproteins in Säugerzellen sorgt, aus der Zelle in das Kulturmedium sekretiert werden. Bevorzugt befinden sich codierte Prozessierungsstellen zwischen dem Leaderfragment und dem Fremdprotein, die entweder in vivo oder in vitro gespalten werden können. Das Leadersequenz-Fragment codiert üblicherweise ein aus hydrophoben Aminosäuren zusammengesetztes Signalpeptid, das die Sekretion des Proteins aus der Zelle steuert. Der dreiteilige Adenovirus-Leader ist ein Beispiel für eine Leadersequenz, die für die Sekretion eines Fremdproteins in Säugerzellen sorgt.

[0060] Üblicherweise sind von Säugerzellen erkannte Transkriptions-Terminations- und Polyadenylierungssequenzen regulatorische Bereiche, die sich 3' vom Translations-Stopcodon befinden und daher zusammen mit den Promotorelementen die codierende Sequenz flankieren. Das 3'-Ende der reifen mRNA wird durch Erkennungsstellen-spezifische post-translationale Spaltung und Polyadenylierung gebildet, Birnstiel et al., *Cell* 41 (1985), 349; Proudfoot and

Whitelaw (1988), Termination and 3' end processing of eukaryotic RNA, in: B. D. Hames und D. M. Glover (Hrsg.), *Transcription and splicing*; Proudfoot, *Trends Biochem. Sci.* 14 (1989), 105. Diese Sequenzen steuern die Transkription einer mRNA, die in ein Protein übersetzt werden kann, das von DNA codiert wird. Beispiele für Transkriptions-Terminator/Polyadenylierungssignale schließen die von SV40 abgeleiteten ein, Sambrook et al. (1989), *Molecular Cloning: A Laboratory Manual*.

[0061] Einige Gene können effizienter exprimiert werden, wenn Introns (auch intervenierende Sequenzen genannt) anwesend sind. Manche cDNAs hingegen wurden effizienter durch Vektoren ohne Spleißsignale (auch Spleiß-Donor- und -Akzeptorstellen genannt) exprimiert, siehe z. B. Gething and Sambrook, *Nature* 293 (1981), 620. Introns sind intervenierende, nicht-codierende Sequenzen innerhalb einer codierenden Sequenz, die Spleiß-Donor und -Akzeptorstellen enthalten. Sie werden durch einen "Spleißen" genannten Prozess entfernt, der auf die Polyadenylierung des primären Transkripts folgt, Nevins, *Annu. Rev. Biochem.* 52 (1983), 441; Green, *Annu. Rev. Genet.* 20 (1986), 671; Padgett et al., *Annu. Rev. Biochem.* 55 (1986), 1119; Krainer and Maniatis (1988), RNA splicing, in: B. D. Hames and D. M. Glover (Hrsg.), *Transcription and splicing*.

[0062] Üblicherweise werden die oben beschriebenen Komponenten, die einen Promotor, ein Polyadenylierungssignal und eine Transkription-Terminationssequenz enthalten, in Expressionskonstrukten zusammengefügt. Enhancer, Introns mit funktionellen Spleiß-Donor- und -Akzeptorstellen und Leadersequenzen können, falls gewünscht, ebenfalls in einem Expressionskonstrukt vorhanden sein. Expressionskonstrukte werden oft in einem Replikon, wie einem extrachromosomalen Element (z. B. einem Plasmid), aufrechterhalten, das zur stabilen Aufrechterhaltung in einer Wirtszelle, wie in Säugerzellen oder Bakterien, befähigt ist. Replikationssysteme aus Säugern schließen die von Tier-Viren ein, die in trans arbeitende Faktoren zur Replikation benötigen. Zum Beispiel replizieren sich Plasmide, die die Replikationssysteme von Papovaviren, wie SV40, Gluzman, *Cell* 23 (1981), 175, oder Polyomavirus enthalten, zu extrem hoher Kopienzahl in Gegenwart des entsprechenden viralen T-Antigens. Weitere Beispiele für Replikons aus Säugern schließen die aus bovinem Papillomavirus und aus Epstein-Barr-Virus abgeleiteten ein. Weiterhin kann das Replikon zwei Replikationssysteme haben, wodurch es, zum Beispiel, in Säugerzellen zur Expression und in einer prokaryotischen Wirtszelle zum Clonieren und zur Amplifikation aufrechterhalten werden kann. Beispiele für solche Säuger-Bakterien-Shuttlevektoren schließen pMT2, Kaufman et al., *Mol. Cell. Biol.* 9 (1989), 946 und pHEBO ein, Shimizu et al., *Mol. Cell. Biol.* 6 (1986), 1074.

[0063] Die Transformations-Methode hängt vom zu transformierenden Wirt ab. Verfahren zum Einbringen von heterologen Polynucleotiden in Säugerzellen sind auf dem Fachgebiet bekannt und schließen Dextran-vermittelte Transfektion, Calciumphosphat-Fällung, Polybren-vermittelte Transfektion, Protoplasten-Fusion, Elektroporation, Einschluss des (der) Polynucleotids (Polynucleotide) in Liposomen, und direkte Mikroinjektion der DNA in die Zellkerne ein.

[0064] Säugerzelllinien, die als Wirtszellen zur Expression verfügbar sind, sind auf dem Fachgebiet bekannt und schließen viele immortalisierte Zelllinien ein, die von der American Type Culture Collection (ATCC) erhältlich sind, eingeschlossen, aber nicht beschränkt auf, Eierstockzellen des Chinesischen Hamsters (CHO), HeLa-Zellen, Baby-Hamster-Nierenzellen (BHK), Affennierenzellen (COS), menschliche Zellen aus hepatozellulären Karzinomen (z. B. Hep G2) und eine Anzahl anderer Zelllinien.

ii. Baculovirus-Systeme

[0065] Das Protein-codierende Polynucleotid kann auch in einen geeigneten Insekten-Expressionsvektor eingefügt werden, und ist funktionell mit den Kontrollelementen innerhalb des Vektors verbunden. Zur Konstruktion von Vektoren werden auf dem Fachgebiet bekannte Techniken verwendet.

[0066] Im Allgemeinen schließen die Komponenten des Expressionssystems einen Transfervektor ein, üblicherweise ein bakterielles Plasmid, das sowohl ein Fragment aus dem Baculovirus-Genom als auch eine geeignete Restriktionsstelle zur Insertion des oder der zu exprimierenden heterologen Gens oder Gene enthält; einen Wildtyp-Baculovirus mit einer zu dem Baculovirus-spezifischen Fragment im Transfervektor homologen Sequenz (dies ermöglicht die homologe Rekombination des heterologen Gens in das Baculovirus-Genom) und geeignete Insekten-Wirtszellen und Züchtungsmedien.

[0067] Nach dem Einfügen der Protein-codierenden DNA-Sequenz in den Transfervektor werden der Vektor und das Wildtyp-Virusgenom in eine Insekten-Wirtszelle transfiziert, wo der Vektor und das virale Genom rekombinieren können. Das verpackte rekombinante Virus wird exprimiert und rekombinante Plaques werden identifiziert und aufgereinigt. Material und Methoden für Baculovirus/Insektenzellen-Expressionssysteme sind im Handel in Kit-Form erhältlich, unter anderem von Invitrogen, San Diego CA ("MaxBac" Kit). Diese Techniken sind dem Fachmann im Allgemeinen bekannt und vollständig beschrieben in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), (nachstehend als "Summers and Smith" bezeichnet).

[0068] Vor der Insertion der Protein-codierenden DNA-Sequenz in das Baculovirus-Genom werden die oben beschriebenen Komponenten, die einen Promotor, einen Leader (falls gewünscht), die codierende Sequenz von Interesse und die Transkriptions-Terminationssequenz umfassen, üblicherweise in einem übertragenden Zwischenkonstrukt (Transfervektor) zusammengesetzt. Dieses Konstrukt kann ein einzelnes Gen und funktionell verbundene regulatorische Elemente; mehrere Gene, jedes mit seinem eigenen Satz funktionell verbundener regulatorischer Elemente; oder mehrere Gene, die durch denselben Satz regulatorischer Elemente reguliert werden, enthalten. Übertragende Zwischenkonstrukte werden häufig in einem Replikon, wie einem extrachromosomalen Element (z. B. Plasmide), aufrechterhalten, das zur stabilen Aufrechterhaltung in einem Wirt, wie einem Bakterium, befähigt ist. Das Replikon hat ein Replikationssystem und kann dadurch in einem geeigneten Wirt zur Clonierung und Amplifikation aufrechterhalten werden.

[0069] Der meistbenutzte Transfervektor zum Einbringen fremder Gene in AcNPV ist gegenwärtig pAc373. Viele andere, dem Fachmann bekannte Vektoren sind ebenfalls entworfen worden. Diese schließen, zum Beispiel, pVL985 ein (der das Polyhedrin-Startcodon von ATG zu ATT ändert, und der eine BamHI-Clonierungsstelle 32 Basenpaare stromabwärts des ATT einführt); siehe Luckow and Summers, Virology 17 (1989), 31.

[0070] Das Plasmid enthält üblicherweise auch das Polyhedrin-Polyadenylierungssignal (Miller et al., Ann. Rev. Microbiol. 42 (1988), 177) und ein prokaryotisches Ampicillin-Resistenzgen (amp) und einen Replikationsursprung zur Selektion und Vermehrung in E. coli.

[0071] Baculovirus-Transfervektoren enthalten üblicherweise einen Baculovirus-Promotor. Ein Baculovirus-Promotor ist jede DNA-Sequenz, die in der Lage ist, eine Baculovirus-RNA-Polymerase zu binden und stromabwärts (5' nach 3') die Transkription einer codierenden Sequenz (z. B. eines Strukturgens) in mRNA zu initiieren. Ein Promotor hat eine Transkriptions-Initiationsregion, die üblicherweise nahe dem 5'-Ende der codierenden Sequenz liegt. Diese Transkriptions-Initiationsregion schließt üblicherweise eine RNA-Polymerase-Bindungsstelle und eine Transkriptions-Initiationsstelle ein. Ein Baculovirus-Transfervektor kann auch eine zweite, Enhancer genannte Domäne besitzen, die, wenn vorhanden, üblicherweise distal zum Strukturgen liegt. Die Expression kann entweder reguliert oder konstitutiv sein.

[0072] Strukturgene, die in den späten Phasen des viralen Infektionszyklus in großer Menge transkribiert werden, liefern besonders nützliche Promotorsequenzen. Beispiele schließen von dem Gen für das

virale Polyhedrin-Protein abgeleitete Sequenzen, Friesen et al. (1986), *The Regulation of Baculovirus Gene Expression*, in: Walter Doerfler (Hrsg.), *The Molecular Biology of Baculoviruses*; EPO-Veröffentlichungen Nr. 127 839 und 155 476; und das das p10-Protein codierende Gen, Vlak et al., *J. Gen. Virol.* 69 (1988), 765, ein.

[0073] Geeignete Signalsequenzen codierende DNA kann aus Genen für sekretierte Insekten- oder Baculovirus-Proteine, wie dem Polyhedrin-Gen von Baculovirus, erhalten werden (Carbonell et al., *Gene* 73 (1988), 409). Da die Signale für posttranslationale Modifikationen in Säugerzellen (wie die Abspaltung des Signalpeptids, proteolytische Spaltung und Phosphorylierung) von Insektenzellen offenbar erkannt werden, und die zur Sekretion und Kernlokalisierung erforderlichen Signale ebenso zwischen den Zellen von Invertebraten und Vertebraten offenbar konserviert sind, können Leadersequenzen nicht-insektoiden Ursprungs, wie die von Genen abgeleiteten, die das humane α -Interferon, Maeda et al., *Nature* 315 (1985), 592; das humane Gastrin-freisetzende Peptid, Lebacqz-Verheyden et al., *Mol. Cell. Biol.* 8 (1988), 3129; das humane IL-2, Smith et al., *Proc. Natl. Acad. Sci. USA* 82 (1985), 8404; das IL-3 der Maus, Miyajima et al., *Gene* 58 (1987), 273; und die humane Glucocerebrosidase, Martin et al., *DNA* 7 (1988), 99, codieren, in einer anderen Ausführungsform auch verwendet werden, um die Sekretion in Insekten zu ermöglichen.

[0074] Ein rekombinantes Polypeptid oder Polyprotein kann intrazellulär exprimiert oder sekretiert werden, wenn es mit geeigneten regulatorischen Sequenzen exprimiert wird. Gute intrazelluläre Expression von nicht-fusionierten Fremdproteinen erfordert üblicherweise heterologe Gene, die idealerweise eine kurze Leadersequenz haben, die geeignete Translations-Initiationssignale vor einem ATG-Startsignal enthält. Falls erforderlich, kann das Methionin am N-Terminus durch in vitro-Inkubation mit Bromcyan vom reifen Protein abgespalten werden.

[0075] In einer anderen Alternative können rekombinante Polyproteine oder Proteine, die natürlicherweise nicht sekretiert werden, durch die Erzeugung chimärer DNA-Moleküle, die für ein Fusionsprotein codieren, das ein Leadersequenz-Fragment enthält, das die Sekretion des Fremdproteins in Insekten ermöglicht, aus der Insektenzelle sekretiert werden. Das Leadersequenz-Fragment codiert üblicherweise ein Signalpeptid, das hydrophobe Aminosäuren enthält, die die Translokation des Proteins in das endoplasmatische Retikulum steuern.

[0076] Nach der Insertion der DNA-Sequenz und/oder des Gens, das das Vorläufer-Expressionsprodukt des Proteins codiert, wird eine Insekten-Wirtszelle mit der heterologen DNA des Transfer-

vektors und der genomischen DNA des Wildtyp-Baculovirus kotransformiert – üblicherweise durch Kointegration. Der Promotor und die Transkriptions-Terminationssequenz des Konstruktes enthält üblicherweise einen 2–5 kb großen Abschnitt des Baculovirus-Genoms. Verfahren zum Einbringen heterologer DNA an die gewünschte Stelle im Baculovirus sind auf dem Fachgebiet bekannt. (Siehe Summers and Smith; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* 3 (1983), 2156 und Luckow and Summers (1989)). Zum Beispiel kann die Insertion in ein Gen wie das Polyhedrin-Gen durch homologe doppelte Crossover-Rekombination erfolgen; die Insertion kann auch in die Erkennungsstelle eines Restriktionsenzym erfolgen, die in das gewünschte Baculovirus-Gen eingeführt wurde. Miller et al., *Bioessays* 4 (1989), 91.

[0077] Die DNA-Sequenz wird sowohl 5' als auch 3' von Polyhedrin-spezifischen Sequenzen flankiert, wenn sie anstelle des Polyhedrin-Gens in den Expressionsvektor cloniert wird, und ist stromabwärts des Polyhedrin-Promotors positioniert.

[0078] Der neu gebildete Baculovirus-Expressionsvektor wird anschließend in ein infektiöses, rekombinantes Baculovirus verpackt. Die homologe Rekombination erfolgt mit niedriger Häufigkeit (zwischen etwa 1% und etwa 5%); daher ist der überwiegende Teil des nach der Kointegration produzierten Virus noch Wildtyp-Virus. Daher ist eine Methode zur Identifizierung rekombinanter Viren erforderlich. Ein Vorteil des Expressionssystems besteht in der visuellen Durchmusterung, die die Unterscheidung rekombinanter Viren gestattet. Das Polyhedrin-Protein, das vom nativen Virus produziert wird, wird in den Zellkernen infizierter Zellen in sehr großen Mengen in den späten Phasen nach der viralen Infektion produziert. Angesammeltes Polyhedrin-Protein bildet Occlusion-Bodies, die auch eingebettete Partikel enthalten. Diese bis zu 15 μ m großen Occlusion-Bodies sind stark lichtbrechend, was ihnen ein hell leuchtendes Aussehen verleiht, das unter dem Lichtmikroskop leicht sichtbar gemacht werden kann. Mit rekombinanten Viren infizierte Zellen besitzen keine Occlusion-Bodies. Um rekombinante Viren von Wildtyp-Viren zu unterscheiden, wird der Transfektionsüberstand durch dem Fachmann bekannte Techniken zur Plaquebildung auf eine einzellige Schicht aus Insektenzellen ausgebracht. Die Plaques werden dann unter dem Lichtmikroskop auf die Anwesenheit (kennzeichnend für Wildtyp-Virus) oder die Abwesenheit (kennzeichnend für rekombinantes Virus) von Occlusion-Bodies durchsucht. Ausubel et al. (Hrsg.) (1990), *Current Protocols in Microbiology*, Band 2, Suppl. 10, Kap. 16.8; Summers and Smith; Miller et al. (1989).

[0079] Rekombinante Baculovirus-Expressionsvektoren wurden für die Infektion von Zellen mehrerer Insektenarten entwickelt. Zum Beispiel wurden rekomb-

binante Baculoviren unter anderem entwickelt für: *Aedes Aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda* und *Trichoplusia ni* (PCT-Veröffentlichung Nr. WO 89/046699; Carbonell et al., J. Virol. 56 (1986), 153; Wright, Nature 312 (1986), 718; Smith et al., Mol. Cell. Biol. 3 (1983), 2156; und siehe allgemein, Fraser et al., In vitro Cell. Dev. Biol. 25 (1989), 225).

[0080] Zellen und Zellkulturmedien sind sowohl für die direkte als auch für die Fusionsexpression heterologer Polypeptide in einem Baculovirus-Expressionssystem im Handel erhältlich; Zellkulturtechnologie ist dem Fachmann im Allgemeinen bekannt. Siehe z. B. Summers and Smith.

[0081] Die veränderten Insektenzellen können dann in einem entsprechenden Nährmedium gezüchtet werden, das eine stabile Aufrechterhaltung des (der) in den veränderten Insekten-Wirtszellen anwesenden Plasmids (Plasmide) erlaubt. Steht das Gen für das Expressionsprodukt unter induzierbarer Kontrolle, können die Wirtszellen zu hohen Dichten gezüchtet und die Expression induziert werden. Wenn in einer anderen Ausführungsform die Expression konstitutiv ist, wird das Produkt kontinuierlich in das Medium exprimiert, und das Nährmedium muss ständig zirkulieren, wobei das gewünschte Produkt abgetrennt und verbrauchte Nährstoffe ergänzt werden. Das Produkt kann durch Techniken wie Chromatographie, z. B. HPLC, Affinitätschromatographie, Ionenaustauschchromatographie, usw.; Elektrophorese; Dichtegradientenzentrifugation; Lösungsmittelextraktion, oder ähnliche gereinigt werden. Falls zweckmäßig, kann das Produkt wie erforderlich weiter gereinigt werden, um im Wesentlichen alle Insektenproteine zu entfernen, die ebenfalls in das Medium sekretiert werden oder durch Lyse von Insektenzellen entstehen, um ein Produkt zu liefern, das zumindest überwiegend frei ist von Bruchstücken der Wirtszellen, wie z. B. Proteinen, Lipiden und Polysacchariden.

[0082] Um Proteinexpression zu erhalten, werden aus den Transformanten abgeleitete rekombinante Wirtszellen unter Bedingungen inkubiert, die die Expression der Sequenz erlauben, die das rekombinante Protein codiert. Diese Bedingungen variieren in Abhängigkeit von der ausgewählten Wirtszelle. Die Bedingungen sind jedoch auf Basis des bekannten Fachwissens für den Durchschnittsfachmann leicht zu ermitteln.

iii. Bakterielle Systeme

[0083] Bakterielle Expressionstechniken sind auf dem Fachgebiet bekannt. Ein bakterieller Promotor ist jede DNA-Sequenz, die in der Lage ist, eine bakterielle RNA-Polymerase zu binden und stromabwärts (3') die Transkription einer codierenden Se-

quenz (z. B. eines Strukturgens) in mRNA zu initiieren. Ein Promotor besitzt eine Transkriptions-Initiationsregion, die sich üblicherweise in der Nähe des 5'-Endes der codierenden Sequenz befindet. Diese Transkriptions-Initiationsregion schließt üblicherweise eine Bindungsstelle für die RNA-Polymerase und eine Transkriptions-Initiationsstelle ein. Ein bakterieller Promotor kann auch eine zweite, Operator genannte Domäne haben, die eine benachbarte Transkriptions-Initiationsstelle überlappen kann, an der die RNA-Synthese beginnt. Der Operator gestattet eine negativ regulierte (induzierbare) Transkription, da das Gen-Repressorprotein an den Operator binden und dadurch die Transkription eines spezifischen Gens inhibieren kann. Konstitutive Expression kann in der Abwesenheit negativ regulatorischer Elemente, wie des Operators, stattfinden. Zusätzlich kann eine positive Regulation durch eine Gen-Aktivatorprotein bindende Sequenz erzielt werden, die sich, falls vorhanden, üblicherweise proximal (5') von der die RNA-Polymerase bindenden Sequenz befindet. Ein Beispiel für ein Gen-Aktivatorprotein ist das Catabolite Activator-Protein (CAP), das bei der Initiierung der Transkription des lac-Operons in *E. coli* mitwirkt, Raibaud et al., Annu. Rev. Genet. 18 (1984), 173. Regulierte Expression kann daher entweder positiv oder negativ sein und dadurch die Transkription entweder verstärken oder abschwächen.

[0084] Sequenzen, die für Enzyme von Stoffwechselwegen codieren, liefern besonders nützlich Promotorsequenzen. Beispiele schließen Promotorsequenzen ein, die sich von Enzymen ableiten, die Zucker wie Galactose, Lactose (lac), Chang et al., Nature 198 (1977), 1056, und Maltose metabolisieren. Weitere Beispiele schließen Promotorsequenzen von biosynthetischen Proteinen ein, wie Tryptophan (trp), Goeddel et al., Nuc. Acids Res. 8 (1980), 4057; Yelverton et al., Nuc. Acids Res. 9 (1981), 731; U.S. 4,738,921; EPO-Veröffentlichungen Nr. 036 776 und 121 775. Das β -Lactamase (bla) Promotor-System, Weissmann (1981), The cloning of interferon and other mistakes., in: I. Gresser (Hrsg.), Interferon 3, das Bakteriophage Lambda p_L -, Shimatake et al., Nature 292 (1981), 128, und das T5-Promotorsystem, U.S. 4,689,406, liefern ebenfalls nützliche Promotorsequenzen.

[0085] Zusätzlich funktionieren auch synthetische Promotoren, die in der Natur nicht vorkommen, als bakterielle Promotoren. Zum Beispiel können Transkriptions-aktivierende Sequenzen eines Bakterien- oder Bakteriophagen-Promotors mit den Operonsequenzen eines anderen Bakterien- oder Bakteriophagen-Promotors verbunden werden, um einen synthetischen Hybridpromotor zu erzeugen, U.S. 4,551,433. Der tac-Promotor zum Beispiel ist ein trp-lac-Hybridpromotor, der sowohl trp-Promotor- als auch lac-Operonsequenzen enthält, und durch den lac-Repressor reguliert wird, Amann et al., Gene 25

(1983), 167; de Boer et al., Proc. Natl. Acad. Sci. USA 80 (1983), 21. Weiterhin kann ein bakterieller Promotor natürlich vorkommende Promotoren nicht-bakteriellen Ursprungs enthalten, die in der Lage sind, eine bakterielle RNA-Polymerase zu binden und die Transkription zu initiieren. Ein natürlich vorkommender Promotor nicht-bakteriellen Ursprungs kann auch mit einer kompatiblen RNA-Polymerase gekoppelt werden, um hohe Expressionsniveaus einiger Gene in Prokaryoten zu erreichen. Das RNA-Polymerase/Promotor-System des Bakteriophagen T7 ist ein Beispiel eines gekoppelten Promotorsystems, Studier et al., J. Mol. Biol. 189 (1986), 113; Tabor et al., Proc. Natl. Acad. Sci. USA 82 (1985), 1074. Darüberhinaus kann ein Hybridpromotor einen Bakteriophagen-Promotor und einen E. coli-Operatorbereich enthalten (EPO-Veröffentlichung Nr. 267 851).

[0086] Zusätzlich zu einer funktionellen Promotorsequenz ist eine effiziente Ribosomen-Bindungsstelle für die Expression von Fremdgenen in Prokaryoten nützlich. In E. coli wird die Ribosomen-Bindungsstelle als Shine-Dalgarno-Sequenz bezeichnet und schließt ein Initiationscodon (ATG) und eine 3–9 Nucleotide lange Sequenz in einem Abstand von 3–11 Nucleotiden stromaufwärts des Initiationscodons ein, Shine et al., Nature 254 (1975), 34. Von der SD-Region wird angenommen, dass sie die Bindung der mRNA an das Ribosom durch Basenpaarung zwischen der SD-Sequenz und dem 3'-Ende der E. coli 16S rRNA begünstigt, Steitz et al. (1979), Genetic signals and nucleotide sequences in messenger RNA., in: R. F. Goldberger (Hrsg.), Biological Regulation and Development: Gene Expression. Um eukaryotische Gene und prokaryotische Gene mit schwachen Ribosom-Bindungsstellen zu exprimieren, Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual.

[0087] Ein DNA-Molekül kann intrazellulär exprimiert werden. Eine Promotorsequenz kann direkt mit dem DNA-Molekül verbunden sein, wobei dann die erste Aminosäure am N-Terminus immer ein Methionin ist, das vom ATG-Startcodon codiert wird. Falls gewünscht, kann das Methionin am N-Terminus vom Protein durch in vitro-Inkubation mit Bromcyan oder entweder durch in vivo- oder in vitro-Inkubation mit einer bakteriellen N-terminalen Methionin-Peptidase abgespalten werden (EPO-Veröffentlichung Nr. 219 237).

[0088] Fusionsproteine bilden eine Alternative zu direkter Expression. Üblicherweise wird eine DNA-Sequenz, die den N-terminalen Teil eines endogenen Bakterienproteins, oder ein anderes stabiles Protein codiert, an das 5'-Ende der heterologen codierenden Sequenz fusioniert. Nach der Expression liefert dieses Konstrukt eine Fusion von zwei Aminosäuresequenzen. Zum Beispiel kann das Zell-Gen des Bak-

teriophagen lambda mit dem 5'-Ende eines Fremdgens verbunden und in Bakterien exprimiert werden. Das entstandene Fusionsprotein beinhaltet bevorzugt eine Erkennungsstelle für ein prozessierendes Enzym (Faktor Xa), um das Bakteriophagen-Protein vom Fremdprotein abzuspalten, Nagai et al., Nature 309 (1984), 810. Fusionsproteine können auch mit Sequenzen des lacZ-Gens, Jia et al., Gene 60 (1987), 197, des trpE-Gens, Allen et al., J. Biotechnol. 5 (1987), 93; Makoff et al., J. Gen. Microbiol. 135 (1989), 11, und EPO-Veröffentlichung Nr. 324 647, hergestellt werden. Die DNA-Sequenz an der Verbindungsstelle der beiden Aminosäuresequenzen kann eine spaltbare oder nicht spaltbare Stelle codieren. Ein anderes Beispiel ist ein Ubiquitin-Fusionsprotein. Ein solches Fusionsprotein wird unter Verwendung der Ubiquitin-Region hergestellt, die bevorzugt eine Erkennungsstelle für ein prozessierendes Enzym beinhaltet (z. B. eine Ubiquitin-spezifische prozessierende Protease), um das Ubiquitin von dem Fremdprotein abzuspalten. Durch dieses Verfahren können native Fremdproteine isoliert werden. Miller et al., Bio/Technology 7 (1989), 698.

[0089] In einer anderen Alternative können Fremdproteine auch von der Zelle durch Erzeugung eines chimären DNA-Moleküls sekretiert werden, das ein Fusionsprotein codiert, das ein Signalpeptidsequenz-Fragment enthält, das die Sekretion des Fremdproteins in Bakterien ermöglicht, U.S. 4,336,336. Das Signalsequenz-Fragment codiert üblicherweise ein Signalpeptid, das hydrophobe Aminosäuren enthält, die die Sekretion des Proteins aus der Zelle steuern. Das Protein wird entweder in das Kulturmedium (Gram-positive Bakterien) oder in das Periplasma, das sich zwischen innerer und äußerer Zellmembran befindet (Gramnegative Bakterien), sekretiert. Bevorzugterweise sind Prozessierungsstellen zwischen dem Signalpeptidfragment und dem Fremdgen codiert, die entweder in vivo oder in vitro gespalten werden können.

[0090] DNA, die geeignete Signalsequenzen codiert, kann aus Genen für sekretierte Bakterienproteine erhalten werden, wie aus dem Gen für das äußere Membranprotein (ompA) von E. coli, Masui et al. (1983), in: Experimental Manipulation of Gene Expression; Ghayeb et al., EMBO J. 3 (1984), 2437, und die Signalsequenz für die Alkalische Phosphatase (phoA) von E. coli, Oka et al., Proc. Natl. Acad. Sci. USA 82 (1985), 7212. Als zusätzliches Beispiel kann die Signalsequenz des alpha-Amylase-Gens verschiedener Bacillus-Stämme zur Sekretion heterologer Proteine aus B. subtilis verwendet werden. Palva et al., Proc. Natl. Acad. Sci. USA 79 (1982), 5582; EPO-Veröffentlichung Nr. 244 042.

[0091] Üblicherweise sind von Bakterien erkannte Transkriptions-Terminationssequenzen regulatorische Regionen, die 3' vom Translations-Stopcodon

liegen, und daher mit dem Promotor die codierende Sequenz flankieren. Diese Sequenzen steuern die Transkription einer mRNA, die in das durch die DNA codierte Polypeptid translatiert werden kann. Transkriptions-Terminationssequenzen beinhalten häufig etwa 50 Nucleotide lange DNA-Sequenzen, die Stamm-Schleifen-Strukturen ausbilden können, die zur Termination der Transkription beitragen. Beispiele schließen Transkriptions-Terminationssequenzen ein, die aus Genen mit starken Promotoren erhalten wurden, wie den *trp*-Genen von *E. coli* und auch anderen biosynthetischen Genen.

[0092] Üblicherweise werden die oben beschriebenen Komponenten, die einen Promotor, eine Signalsequenz (wenn gewünscht), eine bestimmte codierende Sequenz, und eine Transkriptions-Terminationssequenz beinhalten, zu einem Expressionskonstrukt vereinigt. Expressionskonstrukte werden oft in einem Replikon, wie einem extrachromosomalen Element (z. B. Plasmide), aufrechterhalten, das in einem Wirt, wie einem Bakterium, stabil aufrechterhalten wird. Das Replikon besitzt ein replikatives System, das es ihm erlaubt, in einem prokaryotischen Wirt entweder zur Expression oder zur Clonierung und Amplifikation aufrechterhalten zu werden. Weiterhin kann ein Replikon entweder ein Plasmid mit hoher oder niedriger Kopienzahl sein. Ein Plasmid mit hoher Kopienzahl hat im Allgemeinen eine hohe Kopienzahl zwischen etwa 5 bis etwa 200, und üblicherweise etwa 10 bis etwa 150. Eine Wirtszelle, die ein Plasmid mit hoher Kopienzahl enthält, enthält bevorzugt mindestens etwa 10, und stärker bevorzugt mindestens etwa 20 Plasmide. Es kann entweder ein Vektor mit hoher oder niedriger Kopienzahl ausgewählt werden, in Abhängigkeit von der Wirkung des Vektors und des Fremdproteins auf die Wirtszelle.

[0093] In einer anderen Alternative kann das Expressionskonstrukt mit einem integrierenden Vektor in das Bakteriengenom integriert werden. Integrierende Vektoren enthalten üblicherweise mindestens eine zum Bakterienchromosom homologe Sequenz, durch die der Vektor integrieren kann. Integrationen scheinen durch Rekombinationen zwischen homologer DNA im Vektor und dem Bakterienchromosom zu entstehen. Zum Beispiel integrieren sich mit der DNA verschiedener *Bacillus*-Stämme konstruierte integrierende Vektoren in das *Bacillus*-Chromosom (EPO-Veröffentlichung Nr. 127 328). Integrierende Vektoren können auch Bakteriophagen- oder Transposonsequenzen enthalten.

[0094] Üblicherweise können extrachromosomale und integrierende Expressionskonstrukte selektierbare Marker enthalten, um die Selektion transformierter Bakterienstämme zu ermöglichen. Selektierbare Marker können in der bakteriellen Wirtszelle exprimiert werden und können Gene einschließen, die Bakterien resistent gegenüber Antibiotika machen,

wie etwa Ampicillin, Chloramphenicol, Erythromycin, Kanamycin (Neomycin), und Tetracyclin. Davies et al., *Annu. Rev. Microbiol.* 31 (1978), 469. Selektierbare Marker können auch biosynthetische Gene einschließen, wie die des Histidin-, Tryptophan-, und Leucin-Biosynthesewegs.

[0095] In einer anderen Alternative können einige der oben beschriebenen Komponenten in Transformationsvektoren zusammengestellt werden. Transformationsvektoren enthalten üblicherweise einen selektierbaren Marker, der entweder in einem Replikon aufrechterhalten oder in einen integrierenden Vektor entwickelt wird.

[0096] Expressions- und Transformationsvektoren, entweder als extrachromosomale Replikons oder integrierende Vektoren, wurden zur Transformation zahlreicher Bakterienarten entwickelt. Zum Beispiel wurden Expressionsvektoren, unter anderem, für folgende Bakterien entwickelt: *Bacillus subtilis*, Palva et al., *Proc. Natl. Acad. Sci. USA* 79 (1982), 5582; EPO-Veröffentlichungen Nr. 036 259 und 063 953; PCT-Veröffentlichung Nr. WO 84/04541; *E. coli*, Shimatake et al., *Nature* 292 (1981), 128; Amann et al., *Gene* 40 (1985), 183; Studier et al. *J. Mol. Biol.* 189 (1986), 113; EPO-Veröffentlichungen Nr. 036 776, 136 829 und 136 907; *Streptococcus cremoris*, Powell et al., *Appl. Environ. Microbiol.* 54 (1988), 655; *Streptococcus lividans*, Powell et al., *Appl. Environ. Microbiol.* 54 (1988), 655; und *Streptomyces lividans*, U.S. 4,745,056.

[0097] Verfahren zum Einbringen exogener DNA in bakterielle Wirtszellen sind auf dem Fachgebiet wohlbekannt, und schließen üblicherweise die Transformation mit CaCl_2 oder anderen Mitteln, wie divalenten Kationen und DMSO, behandelter Bakterien ein. DNA kann auch durch Elektroporation in Bakterienzellen eingebracht werden. Das Transformationsverfahren ändert sich mit der zu transformierenden Bakterienspezies. Siehe z. B. Masson et al., *FEMS Microbiol. Lett.* 60 (1989), 273; Palva et al., *Proc. Natl. Acad. Sci. USA* 79 (1982), 5582; EPO-Veröffentlichungen Nr. 036 259 und 063 953; PCT-Veröffentlichungen Nr. WO 84/04541, für *Bacillus*; Miller et al., *Proc. Natl. Acad. Sci. USA* 85 (1988), 856; Wang et al., *J. Bacteriol.* 172 (1990), 949, für *Campylobacter*; Cohen et al. *Proc. Natl. Acad. Sci. USA* 69 (1973), 2110; Dower et al. *Nucleic Acids Res.* 16 (1988), 6127; Kushner (1978), An improved method for transformation of *E. coli* with ColE1-derived plasmids, in: H. W. Boyer and S. Nicosia (Hrsg.), *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering*; Mandel et al., *J. Mol. Biol.* 53 (1970), 159; Taketo, *Biochim. Biophys. Acta* 949 (1988), 318, für *Escherichia*; Chassy et al., *FEMS Microbiol. Lett.* 44 (1987), 173, für *Lactobacillus*; Fiedler et al., *Anal. Biochem.* 170 (1988), 38, für *Pseudomonas*; Augustin et al., *FEMS Microbiol. Lett.* 66 (1990),

203, für *Staphylococcus*; Barany et al., J. Bacteriol. 144 (1980), 698; Harlander (1987), Transformation of *Streptococcus lactis* by electroporation, in: J. Ferretti and R. Curtiss III (Hrsg.), *Streptococcal Genetics*; Perry et al., *Infect. Immun.* 32 (1981), 1295; Powell et al., *Appl. Environ. Microbiol.* 54 (1988), 655; Somkuti et al., *Proc. 4th Eur. Cong. Biotechnology* 1 (1987), 412, für *Streptococcus*.

iv. Expression in Hefe

[0098] Hefeexpressionssysteme sind dem Durchschnittsfachmann ebenfalls bekannt. Ein Hefepromotor ist jede DNA-Sequenz, die in der Lage ist, eine Hefe-RNA-Polymerase zu binden und stromabwärts (3') die Transkription einer codierenden Sequenz (z. B. eines Strukturgens) in mRNA zu initiieren. Ein Promotor besitzt eine Transkriptions-Initiationsregion, die sich üblicherweise in der Nähe des 5'-Endes der codierenden Sequenz befindet. Diese Transkriptions-Initiationsregion schließt üblicherweise eine Bindungsstelle für die RNA-Polymerase (die "TA-TA-Box") und eine Transkriptions-Initiationsstelle ein. Ein Hefepromotor kann auch eine zweite, Upstream-Activator-Sequenz (UAS) genannte Domäne haben, die sich, wenn vorhanden, üblicherweise distal zum Strukturgen befindet. Die UAS gestattet eine regulierte (induzierbare) Expression. Konstitutive Expression findet in der Abwesenheit einer UAS statt. Regulierte Expression kann entweder positiv oder negativ sein und dadurch die Transkription entweder verstärken oder abschwächen.

[0099] Hefe ist ein fermentierender Organismus mit einem aktiven Stoffwechselgeschehen, daher liefern Sequenzen, die Enzyme der Stoffwechselwege codieren, besonders nützliche Promotorsequenzen. Beispiele schließen Alkohol-Dehydrogenase (ADH) (EPO-Veröffentlichung Nr. 284 044), Enolase, Glukokinase, Glucose-6-phosphat-Isomerase, Glycerinaldehyd-3-phosphat-Dehydrogenase (GAP oder GAPDH), Hexokinase, Phosphofructokinase, 3-Phosphoglycerat-mutase, und Pyruvatkinase (PyK) (EPO-Veröffentlichung Nr. 329 203) ein. Das PHO5-Gen der Hefe, das eine saure Phosphatase codiert, liefert ebenfalls nützliche Promotorsequenzen, Myanohara et al., *Proc. Natl. Acad. Sci. USA* 80 (1983), 1.

[0100] Darüberhinaus funktionieren synthetische Promotoren, die nicht in der Natur vorkommen, ebenfalls als Hefepromotoren. Zum Beispiel können UAS-Sequenzen eines Hefepromotors mit der Transkriptions-aktivierenden Region eines anderen Hefepromotors verbunden werden, um einen synthetischen Hybridpromotor zu erzeugen. Beispiele für solche Hybridpromotoren schließen die mit der Transkriptions-aktivierenden Region von GAP verbundene regulatorische Sequenz von ADH ein (U.S. 4,876,197 und U.S. 4,880,734). Andere Beispiele für

Hybridpromotoren schließen Promotoren ein, die aus regulatorischen Sequenzen des ADH2-, des GAL4-, des GAL10- oder des PHO5-Gens bestehen, die mit der Transkriptions-aktivierenden Region eines Gens für ein glykolytisches Enzym, wie etwa GAP oder PyK, kombiniert sind (EPO-Veröffentlichung Nr. 164 556). Darüberhinaus kann ein Hefepromotor natürlich vorkommende Promotoren, die nicht aus Hefe stammen, beinhalten, die die Fähigkeit besitzen, die RNA-Polymerase der Hefe zu binden und die Transkription zu initiieren. Beispiele für solche Promotoren schließen, unter anderem, Cohen et al., *Proc. Natl. Acad. Sci. USA* 77 (1980), 1078; Henikoff et al., *Nature* 283 (1981), 835; Hollenberg et al., *Curr. Topics Microbiol. Immunol.* 96 (1981), 119; Hollenberg et al. (1979), *The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae**, in: K. N. Timmis and A. Puhler (Hrsg.), *Plasmids of Medical, Environmental and Commercial Importance*; Mercerau-Puigalon et al., *Gene* 11 (1980), 163; Panthier et al., *Curr. Genet.* 2 (1980), 109, ein.

[0101] Ein DNA-Molekül kann intrazellulär in Hefe exprimiert werden. Eine Promotor-sequenz kann direkt mit dem DNA-Molekül verbunden sein, wobei dann die erste Aminosäure am N-Terminus des rekombinanten Proteins immer ein Methionin ist, das vom ATG-Startcodon codiert wird. Falls gewünscht, kann das Methionin am N-Terminus vom Protein durch in vitro-Inkubation mit Bromcyan abgespalten werden.

[0102] Fusionsproteine stellen eine weitere Möglichkeit für Hefe-Expressionssysteme dar, wie auch in Säuger-, Baculovirus-, und bakteriellen Expressionssystemen. Üblicherweise wird eine DNA-Sequenz, die den N-terminalen Teil eines endogenen Hefeproteins oder ein anderes stabiles Protein codiert, an das 5'-Ende der heterologen codierenden Sequenz fusioniert. Nach der Expression liefert dieses Konstrukt eine Fusion von zwei Aminosäuresequenzen. Zum Beispiel kann das Gen der Superoxid-Dismutase (SOD) aus Hefe oder Mensch mit dem 5'-Ende eines Fremd-Gens verbunden und in Hefe exprimiert werden. Die DNA-Sequenz an der Verbindungsstelle der beiden Aminosäuresequenzen kann eine spaltbare oder nicht spaltbare Stelle codieren. Siehe z. B. EPO-Veröffentlichung Nr. 196 056. Ein anderes Beispiel ist ein Ubiquitin-Fusionsprotein. Ein solches Fusionsprotein wird unter Verwendung der Ubiquitin-Region hergestellt, die bevorzugt eine Erkennungsstelle für ein prozessierendes Enzym beinhaltet (z. B. eine Ubiquitin-spezifische Prozessierungs-Protease), um das Ubiquitin von dem Fremdprotein abzuspalten. Durch dieses Verfahren können daher native Fremdproteine isoliert werden (siehe z. B. PCT-Veröffentlichung Nr. WO 88/024066).

[0103] In einer anderen Alternative können Fremdproteine auch durch Erzeugung chimärer DNA-Mole-

küls, die ein Fusionsprotein codieren, das ein Leadersequenz-Fragment enthält, das die Sekretion des Fremdproteins in Hefe ermöglicht, von der Zelle sekretiert werden. Bevorzugterweise sind Prozessierungsstellen zwischen dem Leaderfragment und dem Fremdgen codiert, die entweder in vivo oder in vitro gespalten werden können. Das Leaderfragment codiert üblicherweise ein Signalpeptid, das hydrophobe Aminosäuren enthält, die die Sekretion des Proteins aus der Zelle steuern.

[0104] DNA, die geeignete Signalsequenzen codiert, kann aus Genen für sekretierte Hefeproteine erhalten werden, wie aus dem Gen der Hefe-Invertase (EPO-Veröffentlichung Nr. 012 873; JPO-Veröffentlichung Nr. 62,096,086) und dem Gen des α -Faktors (U.S. 4,588,684). In einer anderen Ausführungsform gibt es Leader, die nicht aus Hefe stammen, wie etwa ein Interferon-Leader, und ebenfalls die Sekretion in Hefe ermöglichen (EPO-Veröffentlichung Nr. 060 057).

[0105] Eine bevorzugte Klasse von Sekretionsleadern sind solche, die ein Fragment des alpha-Faktor-Gens aus Hefe verwenden, das sowohl eine "Prä"-Signalsequenz als auch eine "Pro"-Region enthält. Die Arten von alpha-Faktor-Fragmenten, die verwendet werden können, schließen sowohl den Voll-Längen-Präpro-alpha-Faktor-Leader (etwa 83 Aminosäurereste) als auch verkürzte alpha-Faktor-Leader (üblicherweise etwa 25 bis etwa 50 Aminosäurereste) ein (U.S. 4,546,083 und U.S. 4,870,008; EPO-Veröffentlichung Nr. 324 274). Zusätzliche Leader, die ein alpha-Faktor-Leaderfragment verwenden, das Sekretion ermöglicht, schließen hybride alpha-Faktor-Leader ein, die aus der Präsequenz des alpha-Faktors einer ersten Hefe, aber mit der pro-Region des alpha-Faktors aus einer zweiten Hefe hergestellt wurden. (Siehe z. B. PCT-Veröffentlichung Nr. WO 89/ 02463).

[0106] Üblicherweise sind von Hefen erkannte Transkriptions-Terminationssequenzen regulatorische Regionen, die 3' vom Translations-Stopcodon liegen, und daher mit dem Promotor die codierende Sequenz flankieren. Diese Sequenzen steuern die Transkription einer mRNA, die in das durch die DNA codierte Polypeptid translatiert werden kann. Beispiele für Transkriptions-Terminationssequenzen und andere von Hefe erkannten Terminationssequenzen, wie solche, die für glykolytische Enzyme codieren.

[0107] Üblicherweise werden die oben beschriebenen Komponenten, die einen Promotor, einen Leader (wenn gewünscht), eine codierende Sequenz von Interesse, und eine Transkriptions-Terminationssequenz beinhalten, zu einem Expressionskonstrukt vereinigt. Expressionskonstrukte werden oft in einem Replikon, wie einem extrachromosomalen Element (z. B. Plasmide), aufrechterhalten, die in einem Wirt, wie einer

Hefe oder einem Bakterium, stabil aufrechterhalten werden. Das Replikon kann zwei Replikationssysteme haben, wodurch es, zum Beispiel, in Hefe zur Expression und in einer prokaryotischen Wirtszelle zum Clonieren und zur Amplifikation, aufrechterhalten werden kann. Beispiele für solche Hefe-Bakterien-Shuttlevektoren schließen YEp24, Botstein et al., Gene 8 (1979), 17-24; pCI/1, Brake et al., Proc. Natl. Acad. Sci. USA 81 (1984), 4642-4646; und YRp17, Stinchcomb et al., J. Mol. Biol. 158 (1982), 157, ein. Weiterhin kann ein Replikon entweder ein Plasmid mit hoher oder niedriger Kopienzahl sein. Ein Plasmid mit hoher Kopienzahl hat im Allgemeinen eine hohe Kopienzahl von etwa 5 bis etwa 200, und üblicherweise von etwa 10 bis etwa 150. Eine Wirtszelle, die ein Plasmid mit hoher Kopienzahl enthält, enthält bevorzugt mindestens etwa 10, und stärker bevorzugt mindestens etwa 20 Plasmide. Es kann entweder ein Vektor mit hoher oder niedriger Kopienzahl ausgewählt werden, in Abhängigkeit der Wirkung des Vektors und des Fremdproteins auf die Wirtszelle.

[0108] In einer anderen Alternative kann das Expressionskonstrukt mit einem integrierenden Vektor in das Hefegenom integriert werden. Integrierende Vektoren enthalten üblicherweise mindestens eine zum Hefechromosom homologe Sequenz, durch die der Vektor integrieren kann, und enthalten bevorzugt zwei homologe Sequenzen, die das Expressionskonstrukt flankieren. Integrationen scheinen durch Rekombinationen zwischen homologer DNA im Vektor und dem Hefechromosom zu entstehen, Orr-Weaver et al., Methods in Enzymol. 101 (1983), 228-245. Ein integrierender Vektor kann durch Auswahl der geeigneten homologen Sequenz als Teil des Vektors zu einem spezifischen Genort in der Hefe gesteuert werden. Eines oder mehrere der Expressionskonstrukte können integrieren, und dadurch die Menge des produzierten rekombinanten Proteins beeinflussen, Rine et al., Proc. Natl. Acad. Sci. USA 80 (1983), 6750. Die im Vektor enthaltenen chromosomalen Sequenzen können entweder als einzelnes Segment im Vektor vorliegen, was zur Integration des gesamten Vektors führt, oder zwei Segmente, die homolog zu aneinander angrenzenden Segmenten im Chromosom sind, flankieren das Expressionskonstrukt im Vektor, was zu einer stabilen Integration des Expressionskonstrukts alleine führen kann.

[0109] Üblicherweise können extrachromosomale und integrierende Expressionskonstrukte selektierbare Marker enthalten, um die Selektion transformierter Hefestämme zu ermöglichen. Selektierbare Marker können biosynthetische Gene einschließen, die in der Hefe-Wirtszelle exprimiert werden können, wie ADE2, HIS4, LEU2, TRP1 und ALG7, und das G418-Resistenzgen, das den Hefezellen Resistenz gegen Tunicamycin bzw. G418 verleiht. Darüberhinaus kann ein geeigneter selektierbarer Marker der Hefe die Fähigkeit zum Wachstum in Anwesenheit to-

xischer Verbindungen, wie Metallen, verleihen. Zum Beispiel ermöglicht die Anwesenheit von CUP1 der Hefe das Wachstum in Gegenwart von Kupferionen. Butt et al., Microbiol. Rev. 51 (1987), 351.

[0110] In einer anderen Alternative können einige der oben beschriebenen Komponenten in Transformationsvektoren zusammengestellt werden. Transformationsvektoren enthalten üblicherweise einen selektierbaren Marker, der entweder in einem Replikon aufrechterhalten oder in einen integrierenden Vektor entwickelt wird.

[0111] Expressions- und Transformationsvektoren, entweder als extrachromosomale Replikons oder integrierende Vektoren, wurden zur Transformation zahlreicher Hefen entwickelt. Zum Beispiel wurden Expressionsvektoren, unter anderem, für folgende Hefen entwickelt: *Candida albicans*, Kurtz et al., Mol. Cell. Biol. 6 (1986), 142; *Candida maltosa*, Kunze et al., J. Basic. Microbiol. 25 (1985), 141; *Hansenula polymorpha*, Gleeson et al., J. Gen. Microbiol. 132 (1986), 3459; Roggenkamp et al., Mol. Gen. Genet. 202 (1986), 302; *Kluyveromyces fragilis*, Das et al., J. Bacteriol. 158 (1984), 1156; *Kluyveromyces lactis*, De Louvencourt et al., J. Bacteriol. 154 (1983), 737; Van den Berg et al., Bio/Technology 8 (1990), 135; *Pichia guilliermondii*, Kunze et al., J. Basic Microbiol. 25 (1985), 141; *Pichia pastoris*, Cregg et al., Mol. Cell. Biol. 5 (1985), 3376; U.S. 4,837,148 und U.S. 4,929,555; *Saccharomyces cerevisiae*, Hinnen et al., Proc. Natl. Acad. Sci. USA 75 (1978), 1929; Ito et al., J. Bacteriol. 153 (1983), 163; *Schizosaccharomyces pombe*, Beach et al., Nature 300 (1981), 706; und *Yarrowia lipolytica*, Davidow et al., Curr. Genet. 10 (1985), 39; Gaillardin et al., Curr. Genet. 10 (1985), 49.

[0112] Verfahren zum Einbringen exogener DNA in Hefe-Wirtszellen sind auf dem Fachgebiet wohlbekannt, und schließen üblicherweise entweder die Transformation von Sphäroplasten oder von mit Alkali-Kationen behandelten Hefezellen ein. Die Transformationsverfahren ändern sich üblicherweise mit der zu transformierenden Hefespezies. Siehe z. B. Kurtz et al., Mol. Cell. Biol. 6 (1986), 142; Kunze et al., J. Basic. Microbiol. 25 (1985), 141, für *Candida*; Gleeson et al., J. Gen. Microbiol. 132 (1986), 3459; Roggenkamp et al., Mol. Gen. Genet. 202 (1986), 302, für *Hansenula*; Das et al., J. Bacteriol. 158 (1984), 1156; De Louvencourt et al., J. Bacteriol. 154 (1983), 737; Van den Berg et al., Bio/Technology 8 (1990), 135, für *Kluyveromyces*; Cregg et al., Mol. Cell. Biol. 5 (1985), 3376; Kunze et al., J. Basic Microbiol. 25 (1985), 141; U.S. 4,837,148 und U.S. 4,929,555, für *Pichia*; Hinnen et al., Proc. Natl. Acad. Sci. USA 75 (1978), 1929; Ito et al., J. Bacteriol. 153 (1983), 163 für *Saccharomyces*; Beach et al., Nature 300 (1981), 706, für *Schizosaccharomyces*; Davidow et al., Curr. Genet. 10 (1985), 39; Gaillardin et al., Curr. Genet. 10

(1985), 49, für *Yarrowia*.

E. Impfstoffe

[0113] Das hierin beschriebene *H. pylori*-Protein kann als alleiniger Kandidat für einen Impfstoff oder in Kombination mit einem oder mehreren Antigenen, letztere entweder von *H. pylori* oder von anderen pathogenen Quellen, verwendet werden. Bevorzugt sind "Cocktail"-Impfstoffe, die zum Beispiel das Cytotoxin-(CT) Antigen, das CAI-Protein und die Uresse enthalten. Zusätzlich kann das hsp zu einer oder mehreren dieser Komponenten hinzugefügt werden. Diese Impfstoffe können entweder prophylaktisch (zur Vorbeugung einer Krankheit) oder therapeutisch (zur Krankheitsbehandlung nach Infektion) sein.

[0114] Diese Impfstoffe enthalten ein *H. pylori*-Antigen oder -Antigene, üblicherweise in Kombination mit "pharmazeutisch verträglichen Trägern", die jeden Träger einschließen, der nicht selbst die Produktion von für das Individuum, das das Mittel erhält, schädlichen Antikörpern verursacht. Geeignete Träger sind typischerweise große, langsam metabolisierte Makromoleküle wie Proteine, Polysaccharide, Polymilchsäuren, Polyglycolsäuren, polymere Aminosäuren, Aminosäure-Copolymere, Lipid-Aggregate (wie etwa Öltröpfchen oder Liposomen), und inaktive Viruspartikel. Diese Träger sind dem Durchschnittsfachmann wohlbekannt. Zusätzlich können diese Träger als immunstimulierende Mittel wirken ("Adjuvantien"). Darüberhinaus kann das Antigen mit einem bakteriellen Toxoid, wie etwa dem Toxoid des Diphtherie-, Tetanus-, Cholera-, *H. pylori*-, und ähnlicher Krankheitserreger versetzt sein.

[0115] Bevorzugte Adjuvantien zur Verbesserung der Wirksamkeit eines Mittels schließen ein, sind aber nicht beschränkt auf: (1) Aluminium-Salze (Alaun), wie etwa Aluminiumhydroxid, Aluminiumphosphat, Aluminiumsulfat, usw.; (2) Öl-in-Wasser-Emulsions-Formulierungen (mit oder ohne andere spezifische immunstimulierende Mittel wie etwa Muramylpeptide (siehe unten) oder Bakterienzellwand-Komponenten), wie zum Beispiel (a) MF59 (PCT-Veröffentlichung Nr. WO 90/14837) mit 5% Squalen, 0.5% Tween 80, und 0.5% Span 85 (das optional verschiedene Mengen MTP-PE (siehe unten) enthält, obgleich nicht erforderlich) in einer Submikron-Partikel-Formulierung, hergestellt unter Verwendung eines Geräts zur Tropfenzerkleinerung, wie etwa Modell 110Y Microfluidizer (Microfluidics, Newton, MA), (b) SAF mit 10% Squalen, 0.4% Tween 80, 5% Pluronic-geblocktem Polymer L121, und thr-MDP (siehe unten), entweder in eine Submicron-Emulsion mikrofluidisiert oder verwirbelt zur Erzeugung einer Emulsion mit größerer Partikelgröße, und (c) RibiTM Adjuvant System (RAS), (Ribi Immunochem, Hamilton, MT) mit 2% Squalen, 0.2% Tween 80, und einer

oder mehreren Bakterienzellwand-Komponenten aus der Gruppe, bestehend aus Monophospholipid A (MPL), Trehalosedimycolat (TDM), und Zellwand-Skelett (CWS), bevorzugt MPL + CWS (Detox™); (3) Saponin-Adjuvantien, wie etwa Stimulon™ (Cambridge Bioscience, Worcester, MA) oder daraus erzeugte Partikel wie etwa ISCOMs (immunstimulierende Komplexe) können verwendet werden; (4) Vollständiges Freund'sches Adjuvans (CFA) und Unvollständiges Freund'sches Adjuvans (IFA); (5) Cytokine, wie Interleukine (IL-1, IL-2, usw.); Makrophagen-Kolonie stimulierender Faktor (M-CSF), Tumornekrosefaktor (TNF), usw.; und (6) andere Substanzen, die als immunstimulierende Mittel zur Verstärkung der Wirksamkeit des Mittels wirken. Alaun und MF59 werden bevorzugt.

[0116] Wie oben erwähnt, schließen Muramylpeptide N-Acetylmuramyl-L-threonyl-D-isoglutamin (thr-MDP), N-Acetyl-normuramyl-L-alanyl-D-isoglutamin (nor-MDP), N-Acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanin-2-(1'-2'-dipalmitoyl-sn-glycerol-3-hydroxyphosphoryloxy)-ethylamin (MTP-PE), usw. ein, sind aber nicht darauf beschränkt.

[0117] Die immunogenen Mittel (z. B. das Antigen, ein pharmazeutisch verträglicher Träger, und ein Adjuvans) enthalten typischerweise Verdünnungsmittel, wie Wasser, Salzlösung, Glycerin, Ethanol, usw. Zusätzlich können Hilfssubstanzen, wie etwa befeuchtende oder emulgierende Mittel, pH-puffernde Substanzen, und ähnliche in diesen Mitteln vorhanden sein.

[0118] Typischerweise werden die immunogenen Mittel injizierbar hergestellt, entweder als flüssige Lösungen oder Suspensionen; feste Formen, die sich zur Lösung oder Suspension in flüssigen Mitteln vor der Injektion eignen, können ebenfalls hergestellt werden. Die Zubereitung kann auch als Emulsion oder eingeschlossen in Liposomen für eine verstärkte Adjuvantien-Wirkung erfolgen, wie oben unter "pharmazeutisch verträgliche Träger" erörtert.

[0119] Immunogene Mittel, die als Impfstoff verwendet werden, beinhalten, wie erforderlich, eine immunologische wirksame Menge des antigenen Polypeptids, sowie eine beliebige andere der oben erwähnten Komponenten, wie notwendig. Mit "immunologisch wirksame Menge" ist gemeint, dass die Verabreichung dieser Menge an ein Individuum, entweder als Einzeldosis oder als Teil einer Serie, wirksam zur Behandlung oder Prävention ist. Diese Menge variiert in Abhängigkeit von Gesundheitszustand und physischer Verfassung des zu behandelnden Individuums, der taxonomischen Gruppe des zu behandelnden Individuums (z. B. Nicht-Mensch-Primat, Primat, usw.), der Kapazität des Immunsystems des Individuums zur Antikörpersynthese, dem gewünschten Grad des Schutzes, der Formulierung des Impfstoffs, der Ein-

schätzung der medizinischen Situation durch den behandelnden Arzt, und anderen relevanten Faktoren. Es ist zu erwarten, dass die Menge in einen relativ breiten Rahmen fällt, der durch Routineversuche bestimmt werden kann.

[0120] Die immunogenen Mittel werden herkömmlich parenteral, z. B. durch entweder subkutane oder intramuskuläre Injektion verabreicht. Weitere Formulierungen, die sich für andere Arten der Verabreichung eignen, schließen orale und pulmonale Formulierungen, Suppositorien und transdermale Applikationen ein. Orale Formulierungen werden für H. pylori-Proteine besonders bevorzugt. Die Dosierung der Behandlung kann nach einem Einzeldosisplan oder einem Mehrfachdosisplan erfolgen. Der Impfstoff kann zusammen mit anderen immunregulatorischen Mitteln verabreicht werden.

F. Immundiagnostische Tests

[0121] H. pylori-Antigene können in Immuntests zum Nachweis von Antikörper-Spiegeln verwendet werden (oder umgekehrt können H. pylori-Antikörper zum Nachweis von Antigen-Spiegeln verwendet werden) und es kann ein Zusammenhang zwischen gastroduodenalen Erkrankungen und besonders duodenalen Geschwüren hergestellt werden. Immuntests auf Basis gut definierter rekombinanter Antigene können entwickelt werden, um die invasiven diagnostischen Verfahren zu ersetzen, die heute angewendet werden. Antikörper gegen H. pylori-Proteine in biologischen Proben, eingeschlossen, zum Beispiel, Blut- oder Serum-Proben, können nachgewiesen werden. Immuntests können auf sehr unterschiedliche Weise aufgebaut werden, und verschiedene Möglichkeiten dazu sind auf dem Fachgebiet bekannt. Protokolle für Immuntests können zum Beispiel auf Konkurrenz, oder direkter Reaktion, oder Tests vom Sandwich-Typ beruhen. Die Protokolle können zum Beispiel auch feste Träger verwenden, oder können mit Immunopräzipitation arbeiten. Die meisten Tests umfassen die Verwendung markierter Antikörper oder Polypeptide; die Markierungen können zum Beispiel fluoreszierende, chemilumineszente, radioaktive oder Farbstoff-Moleküle sein. Tests, die die Signale der Sonde verstärken, sind ebenfalls bekannt; Beispiele dafür sind Tests, die Biotin und Avidin verwenden, und Enzym-markierte und indirekte Immuntests, wie etwa ELISA-Tests.

[0122] Kits, die sich zur Immundiagnostik eignen und die entsprechenden markierten Reagentien enthalten, werden durch Verpacken der entsprechenden Substanzen, einschließlich der erfindungsgemäßen Mittel, in geeignete Behälter zusammengestellt, zusammen mit den anderen Reagentien und Substanzen (zum Beispiel geeigneten Puffern, Salzlösungen, usw.), die zur Testdurchführung erforderlich sind, wie auch einer geeigneten Zusammenstellung von Test-

vorschriften.

G. Beispiele

[0123] Die im Folgenden dargestellten Beispiele sind als weitere Anleitung für den Durchschnittsfachmann bestimmt, und sollen in keinsten Weise als Beschränkung der Erfindung ausgelegt werden.

1. Material und Methoden

a. Clonierung

[0124] Zwei Mischungen degenerierter Oligonucleotide wurden unter Verwendung eines Applied Biosystems Modell 380B DNA-Synthesizers synthetisiert. Diese Mischungen wurden in einer 4-mikromolaren Konzentration in einer 100-Mikroliter-Polymerase-Kettenreaktion mit 200 Nanogramm gereinigter DNA unter Verwendung des Genamp PCR-Kits nach Vorschrift des Herstellers verwendet. Die Reaktion wurde für 1 Minute bei 94°C, 2 Minuten bei 48°C und 2 Minuten bei 56°C inkubiert. Die Reaktionsmischung durchlief 30 Zyklen unter diesen Bedingungen.

[0125] Die Analyse der Reaktionsprodukte durch Agarose-Gelelektrophorese ergab ein markantes DNA-Fragment von ungefähr 87 bp. Nach Spaltung mit den Restriktionsenzymen XbaI und EcoRI wurde das Fragment in das zuvor ebenfalls mit XbaI und EcoRI gespaltene Plasmid Bluescript SK+ (Stratagene) ligiert. Der Ligierungsansatz wurde zur Transformation kompetenter E. coli durch Elektroporation bei 2000 V und 25 Mikrofarad (200 Ω) mit einem BioRad Gene Pulser (Kalifornien) verwendet. Transformierte E. coli wurden durch Anzucht auf L-Agarplatten selektiert, die 100 Mikrogramm Ampicillin pro Milliliter enthielten. Plasmid-DNA wurde aus positiven E. coli-Isolaten extrahiert und einer Sequenzanalyse unter Verwendung des Sequenase 2-DNA-Sequenzierungskits (United States Biochemical Corporation) nach Vorschrift des Herstellers unterzogen.

b. Anlegen von Genbanken

1) Genbank aus HindIII-Fragmenten

[0126] Sieben Mikrogramm gereinigter DNA wurden vollständig mit dem Restriktionsenzym HindIII gespalten. Drei Mikrogramm Bluescript SK+-Plasmid wurden vollständig mit HindIII gespalten, dann mit Phosphatase aus Kälberdarm behandelt. Beide DNA-Gemische wurden durch Ausschütteln mit Wasser-gesättigtem Phenol gereinigt, dann durch Zugabe von Ethylalkohol auf 67% (v/v) gefällt. Beide DNAs wurden in 50 Mikroliter Wasser resuspendiert. 0.7 Mikrogramm der DNA-Fragmente wurden mit 0.3 Mikrogramm der Bluescript-DNA in 50 Mikroliter einer Lösung gemischt, die 25 mM Tris pH 7.5, 10 mM MgCl₂ und 5 Units T4 DNA-Ligase enthielt. Diese Mischung

wurde bei 15°C für 20 Stunden inkubiert, wonach die DNA mit Wasser-gesättigtem Phenol extrahiert und mit Ethylalkohol gefällt wurde. Daraufhin wurde die DNA in 50 Mikrol. Wasser resuspendiert. Die Einführung von 1 Mikrol. dieser DNA in E. coli durch Elektroporation erbrachte ungefähr 3000–10,000 Ampicillin-resistente Bakterienkolonien.

2) Genbank aus EcoRI-Fragmenten

[0127] Etwa 0.7 Mikrog EcoRI-gespaltener DNA wurden gereinigt und mit 0.45 Mikrogramm Bluescript SK+ Plasmid gemischt, das zuvor mit EcoRI gespalten und mit Phosphatase aus Kälberdarm behandelt worden war. Die Fragmente wurden in 50 Mikrol. Lösung ligiert. Nach Reinigung und Fällung wurde die DNA in 50 Mikrol. Wasser resuspendiert. Elektroporation von E. coli mit 1 Mikrol. dieser Lösung erbrachte ungefähr 200 Ampicillin-resistente Bakterienkolonien.

[0128] Um geeignete Restriktionsfragmente aus dem Genom für die weitere Clonierung zu identifizieren, wurde das Plasmid einheitlich mit ³²P markiert und als Sonde verwendet, um DNA des Stammes CCUG zu analysieren, die mit verschiedenen Restriktionsenzymen gespalten, über Agarose-Gelelektrophorese aufgetrennt und auf Nitrocellulose-Filter übertragen wurde. Die Sonde zeigte ein einzelnes HindIII-Restriktionsfragment von ungefähr 3.5 kb. Eine Genbank aus HindIII-gespaltenen DNA-Fragmenten wurde hergestellt und in den Bluescript-Plasmidvektor cloniert. Diese Genbank wurde mit der ³²P-markierten DNA abgesucht, die dem zuvor clonierten 87 bp Fragment entsprach. Zwei Clone, die identische HindIII-Fragmente von ungefähr 3.3 kbp enthielten, wurden identifiziert.

[0129] Die DNA-Sequenzierung dieser HindIII-Fragmente ergab Sequenzen, die die 23 Aminosäuren codieren können, die mit dem Aminoterminal des zuvor beschriebenen 87 kDa-Cytotoxins übereinstimmen. Diese Sequenzen enthalten den Teil eines offenen Leserahmens mit ungefähr 300 Nucleotiden, der am Rand des Fragments endet, das durch die HindIII-Restriktionsstelle begrenzt wird. Die Sequenz zeigte ebenso das Vorhandensein einer EcoRI-Restriktionsstelle innerhalb des vermuteten offenen Leserahmens in 120 bp Entfernung von der HindIII-Stelle.

[0130] Eine ³²P-markierte Sonde, die mit der Sequenz zwischen der EcoRI-Stelle und der HindIII-Stelle übereinstimmte, wurde zum Absuchen einer Genbank aus EcoRI-DNA-Fragmenten, cloniert in den Bluescript SK+ Vektor, verwendet. Diese Sonde ergab zwei Clone, die Fragmente von ungefähr 7.3 kbp enthielten. Die DNA-Sequenzierung dieser Fragmente ergab einen durchgehenden offenen Leserahmen, der sich mit den aus den 3.2 kbp HindIII-Fragmenten bestimmten Sequenzen überschneidet. Die

DNA-Sequenz dieser überlappenden Fragmente und die konzeptionelle Translation des darin enthaltenen einzelnen offenen Leserahmens sind in **Fig. 1** bzw. **Fig. 2** gezeigt.

[0131] Es soll angemerkt werden, dass sich diese Clone als extrem instabil erwiesen. Die beim Absuchen identifizierten ursprünglichen Kolonien waren so klein, dass sie schwierig zu finden waren. Die Vermehrung dieser Clone durch herkömmliche Verfahren der nachfolgenden Züchtung für 16–18 Stunden ergab sehr heterogene Plasmidpopulationen aufgrund von DNA-Umlagerungen und Deletionen. Ausreichende Mengen dieser Clone wurden durch nachfolgende Züchtung für 8–10 Stunden ohne Antibiotika-Selektion angezogen. Obwohl die Plasmidausbeuten relativ gering waren, wurden auf diese Weise Selektion und Wachstum von Bakterien verhindert, die ein vermehrungsfähiges umgelagertes Plasmid enthielten.

c. Absuchen von DNA-Genbanken

[0132] Das Produkt der PCR-Reaktion, das das markante 87 bp-Fragment enthielt, wurde mit ^{32}P durch Random Priming unter Verwendung des Prime-a-gene Kits (Promega) markiert. Diese markierte Sonde wurde in einer Hybridisierungsreaktion mit an Nitrocellulosefiltern immobilisierter DNA von ungefähr 3000 Bakterienclonen verwendet. Die Hybridisierungsreaktion wurde bei 60°C in einer 0.3 M NaCl-Lösung durchgeführt. Ein positiver Bakterienclon wurde vermehrt und die Plasmid-DNA wurde präpariert. Das Plasmid enthielt ein Insert von ungefähr 3.3 kb DNA und wurde TOXHH1 genannt.

[0133] Ein 120 bp-Fragment, das die in **Fig. 1** gezeigte Sequenz zwischen Position 292 und 410 enthielt, wurde aus dem Plasmid TOXHH1 gewonnen und zum Absuchen von ungefähr 400 Kolonien der Genbank aus EcoRI-Fragmenten verwendet. Ein positiver Clon, der ungefähr 7.3 kb DNA-Sequenzen enthielt, wurde isoliert und TOXEE1 genannt.

[0134] Die in **Fig. 1** gezeigte Nucleotidsequenz wurde aus den Clonen TOXHH1 und TOXEE1 unter Verwendung des Sequenase 2-Sequenzierungskits erhalten. Die Nucleotide zwischen Position 1 und 410 in **Fig. 1** wurden aus TOXHH1 und die zwischen 291 und 3507 aus TOXEE1 erhalten. *E. coli*, die die Plasmide TOXHH1 und TOXEE1 enthielten, wurden bei der American Type Culture Collection hinterlegt, siehe unten.

d. Herstellung von Antiseren gegen das Cytotoxin

[0135] Ein den Nucleotiden 116–413 der in **Fig. 1** gezeigten Sequenz entsprechendes DNA-Fragment wurde so in den bakteriellen Expressionsvektor pex 34 A cloniert, dass durch Induktion des Bakterienpro-

motors ein Fusionsprotein produziert wurde, das einen Teil des MS2-Polymerase-Polypeptids enthielt, das mit den Aminosäuren des Cytotoxin-Polypeptids fusioniert war und die zuvor identifizierten 23 Aminosäuren einschloss. Ungefähr 200 Mikrogramm dieses Fusionsproteins wurden durch Acrylamid-Gelelektrophorese teilweise gereinigt und zur Immunisierung von Kaninchen durch Standardverfahren verwendet.

[0136] Antiseren dieser Kaninchen, die nach 3 Immunisierungen im Abstand von je 1 Monat entnommen wurden, wurden zur Untersuchung von Proteinextrakten eines Cytotoxin-positiven und eines Cytotoxin-negativen *H. pylori*-Stammes in üblichen Immunoblot-Experimenten verwendet.

[0137] Die Antiseren reagierten mit einem Polypeptid, das in der denaturierenden Polyacrylamid-Gelelektrophorese mit einem scheinbaren Molekulargewicht von 100 kDa wanderte. Dieses Polypeptid wurde in Proteinextrakten des Cytotoxin-positiven, aber nicht des Cytotoxin-negativen Stammes gefunden. Vor der Immunisierung gewonnenes Serum reagierte nicht mit diesem Polypeptid.

e. Teilweise Aufreinigung der vakuolisierenden Aktivität

[0138] Vollständige *H. pylori*-Membranen wurden in einer Konzentration von 6 mg/ml in einer Lösung mit 1% CHAPS, 0.5 M NaCl, 10 mM Hepes pH 7.4, 2.5 mM EDTA, 20% Saccharose für 1 Stunde bei 4°C solubilisiert. Dieses Gemisch wurde auf einen diskontinuierlichen Saccharose-Gradienten mit Stufen von 30%, 35%, 40% und 55% Saccharose aufgetragen und einer Ultrazentrifugation für 17 Stunden bei 20000 × g unterzogen. Der Gradient wurde fraktioniert und jede Fraktion wurde auf vakuolisierende Aktivität und auf Uresse-Aktivität getestet. Mit der Uresse-Aktivität assoziierte vakuolisierende Aktivität wurde in mehreren Fraktionen des Gradienten gefunden. Ein Signal der vakuolisierenden Aktivität wurde ebenfalls in den obersten Fraktionen des Gradienten gefunden und diese Fraktionen waren praktisch frei von Uresse-Aktivität.

[0139] Diese Uresse-unabhängige vakuolisierende Aktivität wurde durch stufenweise Fällung mit Ammoniumsulfat-Konzentrationen von 20% bis 34% weiter fraktioniert. Denaturierende Polyacrylamid-Gelelektrophorese der bei verschiedenen Ammoniumsulfat-Konzentrationen gefällten Proteine ergab ein markantes Polypeptid von etwa 100 kDa, das zusammen mit der vakuolisierenden Aktivität gereinigt wurde. Dieses Polypeptid wurde von den gegen die oben beschriebenen rekombinanten Fusionsproteine gewonnenen Kaninchen-Antiseren erkannt.

2. Ergebnisse

[0140] Zwei überlappende Fragmente, die etwa 10 kbp des *H. pylori*-Genoms entsprechen, wurden cloniert. Diese Clone enthalten ein Gen mit 3960 bp (gezeigt in **Fig. 1**), das für ein Polypeptid mit 1296 Aminosäuren codieren kann (gezeigt in **Fig. 2**). Das Molekulargewicht dieses mutmaßlichen Polypeptids beträgt 139.8 kDa. Die Nucleotidsequenz AGGAAG 9 bp stromaufwärts des Methionin-Codons an Position 18 in **Fig. 1** ähnelt stark der Shine-Dalgarno-Consensussequenz und unterstützt die Hypothese, dass dieses Methionin das Initiator-Methionin für die Synthese des Polypeptids repräsentiert. Eine 30 bp-Nucleotidsequenz, die 10 bp stromabwärts des vermuteten Stopcodons an Position 3906 in **Fig. 1** beginnt, ähnelt stark der Struktur eines prokaryotischen Transkriptions-Terminators und repräsentiert wahrscheinlich das Ende der codierenden Sequenzen der Boten-RNA.

[0141] Das Cytotoxin-Gen wird durch folgende Kriterien definiert, den Polypeptid-Vorläufer der vakuolisierenden Aktivität von *H. pylori* zu codieren:

- (i) Das mutmaßliche Polypeptid enthält die 23 Aminosäuren lange Sequenz (**Fig. 2**, Positionen 34–56), die als Aminoterminus des zuvor beschriebenen 87 kDa vakuolisierenden Proteins identifiziert wurden, Clover et al., J. Biol. Chem. 267 (1992), 10570–75. Dieser Sequenz gehen 33 Aminosäuren voraus, die einer prokaryotischen Leadersequenz ähneln; daher repräsentiert diese Sequenz wahrscheinlich den Aminoterminus eines reifen Proteins;
- (ii) Kaninchen-Antiseren, die für ein 100 Aminosäuren langes Fragment des mutmaßlichen Polypeptids spezifisch sind, das den vorgeschlagenen Aminoterminus enthält, erkannten ein 100 kDa-Polypeptid in einem Cytotoxin-positiven, nicht hingegen in einem Cytotoxin-negativen *H. pylori*-Stamm. Dieses 100 kDa-Polypeptid wird zusammen mit der vakuolisierenden Aktivität aus *H. pylori*-Membranen aufgereinigt.

[0142] Zusammengenommen codiert das hier beschriebene Gen ein ungefähr 140 kDa-Polypeptid, das zu einem 100 kDa-Polypeptid prozessiert wird, das mit der cytotoxischen Aktivität von *H. pylori* in Zusammenhang steht. Das zuvor beschriebene 87 kDa-Polypeptid muss entweder durch weitere Prozessierung des 100 kDa-Polypeptids oder durch proteolytischen Abbau während der Aufreinigung entstehen.

H. Hinterlegung von biologischem Material

[0143] Das folgende Material wurde bei der American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland, Telefon (301) 231–5519, gemäß den Bedingungen des Budapester Vertrags

zur Internationalen Anerkennung der Hinterlegung von Mikroorganismen zum Zweck des Patentverfahrens hinterlegt.

[0144] Für das cytotoxische Protein (CT): ATCC Nr. 69157 *E. coli* TG1, enthaltend Plasmid TOXHH1

[0145] Diese Hinterlegungen sind zum Vorteil des Fachmanns bereitgestellt. Die Nucleinsäuresequenzen dieser Hinterlegungen, wie auch die Aminosäuresequenzen der dadurch codierten Polypeptide, sollten im Falle eines jeglichen Fehlers in den hier beschriebenen Sequenzen im Vergleich zu den Sequenzen der Hinterlegungen als Bezug genannt werden. Eine Lizenz zur Herstellung, Gebrauch, oder Verkauf des hinterlegten Materials kann erforderlich sein, und keine solche Lizenz wird hiermit bewilligt.

Patentansprüche

1. Protein, ausgewählt aus der Gruppe bestehend aus:

- (a) einem *Helicobacter pylori*-Cytotoxin mit der Aminosäuresequenz wie in **Fig. 2** gezeigt;
- (b) einem *Helicobacter pylori*-Cytotoxin mit der Aminosäuresequenz wie in **Fig. 2** gezeigt, modifiziert durch eine oder mehrere konservative Aminosäureaustausche;
- (c) einem Protein, welches (i) immunologisch identifizierbar ist durch einen Antikörper, welcher spezifisch mit dem *Helicobacter pylori*-Cytotoxin von (a) reagiert und (ii) ein N-terminales Met aufweist;
- (d) einem *Helicobacter pylori*-Cytotoxin, umfassend die Aminosäure-Signalsequenz
MEIQQTHRKINRPLVSLALVGALVSITPQQSHA;
- (e) einem Protein, umfassend (i) ein N-terminales Met und (ii) ein Fragment der Aminosäuresequenz wie in **Fig. 2** gezeigt, wobei das Fragment ein Teil von mindestens 8 Aminosäuren dieser Sequenz ist.

2. Protein nach Anspruch 1, welches markiert oder an einen festen Träger gekoppelt ist.

3. Protein nach Anspruch 1 oder Anspruch 2, zur Verwendung in der Behandlung einer *Helicobacter pylori*-Infektion.

4. Protein nach Anspruch 1 oder Anspruch 2 zur Verwendung als Impfstoff.

5. Impfstoff oder therapeutische Zusammensetzung, umfassend das Protein nach Anspruch 1 oder Anspruch 2 und einen pharmazeutisch verträglichen Träger.

6. Impfstoff oder therapeutische Zusammensetzung nach Anspruch 5, weiterhin umfassend ein oder mehrere der folgenden:

- i) *Helicobacter pylori*-Cytotoxin-assoziiertes immuno-

dominantes Antigen oder Vorläufer davon;
 ii) rekombinantes *Helicobacter pylori*-Hitzeschockprotein;
 iii) *Helicobacter pylori*-Uresse.

7. Impfstoff oder therapeutische Zusammensetzung nach Anspruch 5 oder Anspruch 6, weiterhin umfassend ein Adjuvans.

8. Verfahren zur Herstellung eines Impfstoffs oder einer therapeutischen Zusammensetzung nach einem der Ansprüche 5 bis 7, umfassend das Zusammenbringen eines oder mehrerer Proteine nach Anspruch 1 oder Anspruch 2 mit einem pharmazeutisch verträglichen Träger und, gegebenenfalls, einem Adjuvans.

9. In-vitro-immundiagnostischer Test, umfassend mindestens einen Schritt, der als mindestens einen Bindepartner ein Protein nach Anspruch 1 oder Anspruch 2 beinhaltet, welches gegebenenfalls markiert oder an einen festen Träger gekoppelt ist.

10. Kit zur Immundiagnose zur Durchführung eines Tests nach Anspruch 9, umfassend mindestens ein Protein nach Anspruch 1 oder Anspruch 2.

11. Verwendung eines oder mehrerer Proteins/Proteine nach Anspruch 1 oder 2 für die Herstellung eines Medikaments zur Behandlung oder Verhinderung einer *Helicobacter pylori*-Infektion.

12. Polynucleotid, ausgewählt aus der Gruppe bestehend aus:

- (a) einem Polynucleotid, welches ein Protein nach Anspruch 1 codiert;
- (b) einem Polynucleotid mit der Nucleotidsequenz, die den Nucleotiden 291–3507 der in **Fig. 1** gezeigten Sequenz entspricht;
- (c) einem Polynucleotid, das ein Fragment der Nucleotidsequenz unter (b) umfasst, wobei das Fragment ein Protein von mindestens 8 Aminosäuren codiert.

13. Polynucleotid-Einfangsonde, umfassend mindestens 8 zusammenhängende Nucleotide (a) des Polynucleotids nach Anspruch 12 oder (b) des Komplements von (a).

14. Einfangsonde nach Anspruch 13, umfassend mindestens 14 zusammenhängende Nucleotide von (a) oder (b).

15. Nucleinsäure-Test, wobei in mindestens einem Schritt die Sonde nach Anspruch 13 oder Anspruch 14 verwendet wird.

16. Kit zur Durchführung eines Nucleinsäure-Tests, umfassend mindestens eine Polynucleotid-Sonde nach Anspruch 13 oder 14.

17. Polynucleotid-Amplifikationsverfahren, in dem ein Polynucleotidprimer verwendet wird, wobei mindestens ein Primer ein Polynucleotid ist, welches mindestens 14 zusammenhängende Nucleotide (a) des Polynucleotids nach Anspruch 12 oder (b) des Komplements von (a) umfasst.

18. Kit zur Durchführung eines Polynucleotid-Amplifikationsverfahrens, in dem ein Polynucleotidprimer verwendet wird, wobei mindestens ein Primer ein Polynucleotid ist, welches mindestens 14 zusammenhängende Nucleotide (a) des Polynucleotids nach Anspruch 12 oder (b) des Komplements von (a) umfasst.

19. Vektor, umfassend ein rekombinantes Polynucleotid nach Anspruch 12.

20. Wirtszelle, transformiert mit einem Vektor nach Anspruch 19.

21. Verfahren zur Herstellung eines rekombinanten Polypeptids nach Anspruch 1 oder 2, umfassend das Züchten einer Wirtszelle nach Anspruch 20 und Isolieren des rekombinanten Polypeptids.

Es folgen 14 Blatt Zeichnungen

Anhängende Zeichnungen

1 AAAAAGAAAG GAAGAAAATG GAAATACAAC AAACACACCG CAAAATCAAT
51 CGCCCTCTGG TTTCTCTCGC TTTAGTAGGA GCATTAGTCA GCATCACACC
101 GCAACAAAGT CATGCCGCCT TTTTCACAAC CGTGATCATT CCAGCCATTG
151 TTGGGGGTAT CGCTACAGGC ACCGCTGTAG GAACGGTCTC AGGGCTTCTT
201 AGCTGGGGGC TCAAACAAGC CGAAGAAGCC AATAAAACCC CAGATAAACC
251 CGATAAAGTT TGGCGCATT C AAGCAGGAAA AGGCTTTAAT GAATTCCCTA
301 ACAAGGAATA CGACTTATAC AGATCCCTTT TATCCAGTAA GATTGATGGA
351 GGTGTTGGGATT GGGGGAATGC CGCTAGGCAT TATTGGGTCA AAGGCGGGCA
401 ACAGAATAAG CTTGAAGTGG ATATGAAAGA CGCTGTAGGG ACTTATACCT
451 TATCAGGGCT TAGAACTTT ACTGGTGGGG ATTTAGATGT CAATATGCAA
501 AAAGCCACTT TACGCTTGGG CCAATTCAAT GGCAATTCTT TTACAAGCTA
551 TAAGGATAGT GCTGATCGCA CCACGAGAGT GATTTCAACG CTAAAAATAT
601 CTCAATTGAT AATTTTGCAG AAATCAACAA CTCGTGTGGG TTCTGGAGCC
651 GGGAGGAAAG CCAGCTCTAC GGTTTTGACT TTGCAAGCTT CAGAAGGGAT
701 CACTAGCGAT AAAAACGCTG AAATTTCTCT TTATGATGGT GCCACGCTCA
751 ATTTGGCTTC AAGCAGCGTT AAATTAATGG GTAATGTGTG GATGGGCCGT
801 TTGCAATACG TGGGAGCGTA TTTGGCCCCT TCATACAGCA CGATAAACAC
851 TTCAAAAGTA ACAGGGGAAG TGAATTTTAA CCACCTCACT GTTGGCGATA
901 AAAACGCCGC TCAAGCGGGC ATTATCGCTA ATAAAAAGAC TAATATTGGC
951 AACTGGATT TGTGGCAAAG CGCCGGGTTA AACATTATCG CTCCTCCAGA
1001 AGGTGGCTAT AAGGATAAAC CCAATAATAC CCCTTCTCAA AGTGGTGCTA
1051 AAAACGACAA AAATGAAAGC GCTAAAAACG ACAAACAAGA GAGCAGTCAA
1101 AATAATAGTA AACTCAGGT CATTAAACCA CCAATAGTG CGCAAAAAAC
1151 AGAAGTTCAA CCCACGCAAG TCATTGATGG GCCTTTTGCG GGCGGCAAAG
1201 ACACGGTTGT CAATATCAAC CGCATCAACA CTAACGCTGA TGGCACGATT
1251 AGAGTGGGAG GGTTTAAAGC TTCTCTTACC ACCAATGCGG CTCATTTGCA
1301 TATCGGCAAA GCGGGTGTCA ATCTGTCCAA TCAAGCGAGC GGGCGCTCTC

FIG. 1A

1351 TTATAGTGGA AAATCTAACT GGGAAATATCA CCGTTGATGG GCCTTTAAGA
 1401 GTGAATAATC AAGTGGGTGG CTATGCTTTG GCAGGATCAA GCGCGAATTT
 1451 TGAGTTTAAAG GCTGGTACGG ATACCAAAAA CGGCACAGCC ACTTTTAAATA
 1501 ACGATATTAG TCTGGGAAGA TTTGTGAATT TAAAGGTGGA TGCTCATACA
 1551 GCTAATTTTA AAGGTATTGA TACGGGTAAT GGTGGTTTCA ACACCTTAGA
 1601 TTTTAGTGGC GTTACAGACA AAGTCAATAT CAACAAGCTC ATTACGGCTT
 1651 CCACTAATGT GGCCGTAAA AACTTCAACA TTAATGAATT GATTGTTAAA
 1701 ACCAATGGGA TAAGTGTGGG GGAATATACT CATTTTAGCG AAGATATAGG
 1751 CAGTCAATCG CGCATCAATA CCGTGCCTT GGAAACTGGC ACTAGGTCAC
 1801 TTTTCTCTGG GGGTGTAAA TTAAAGGTG GCGAAAAATT GGTATAGAT
 1851 GAGTTTTACT ATAGCCCTTG GAATTATTTT GACGCTAGAA ATATTAAAAA
 1901 TGTTGAAATC ACCAATAAAC TTGCTTTTGG ACCTCAAGGA AGTCCTTGGG
 1951 GCACATCAAA ACTTATGTTC AATAATCTAA CCCTAGGTCA AAATGCGGTC
 2001 ATGGATTATA GCCAATTTTT AAATTTAACC ATTCAAGGGG ATTTTCATCAA
 2051 CAATCAAGGC ACTATCAACT ATCTGGTCCG AGGTGGGAAA GTGGCAACCT
 2101 TAAGCGTAGG CAATGCAGCA GCTATGATGT TTAATAATGA TATAGACAGC
 2151 GCGACCGGAT TTTACAAACC GCTCATCAAG ATTAACAGCG CTCAAGATCT
 2201 CATTAAAAAT ACAGAACATG TTTTATTGAA AGCGAAAATC ATTGGTTATG
 2251 GTAATGTTTC TACAGGTACC AATGGCATTG GTAATGTAA TCTAGAAGAG
 2301 CAATTCAAAG AGCGCCTAGC CCTTTATAAC AACAATAACC GCATGGATAC
 2351 TTGTGTGGTG CGAAATACTG ATGACATTAA AGCATGCGGT ATGGCTATCG
 2401 GCGATCAAAG CATGGTGAAC AACCTGACA ATTACAAGTA TCTTATCGGT
 2451 AAGGCATGGA AAAATATAGG GATCAGCAAA ACAGCTAATG GCTCTAAAAT
 2501 TTCGGTGTAT TATTTAGGCA ATTCTACGCC TACTGAGAAT GGTGGCAATA
 2551 CCACAAATTT ACCCACAAC ACCACTAGCA ATGCACGTTC TGCCAACAAC
 2601 GCCCTTGCAC AAAACGCTCC TTTCGCTCAA CCTAGTGCTA CTCCTAATTT
 2651 AGTCGCTATC AATCAGCATG ATTTTGGCAC TATTGAAAGC GTGTTTGAAT

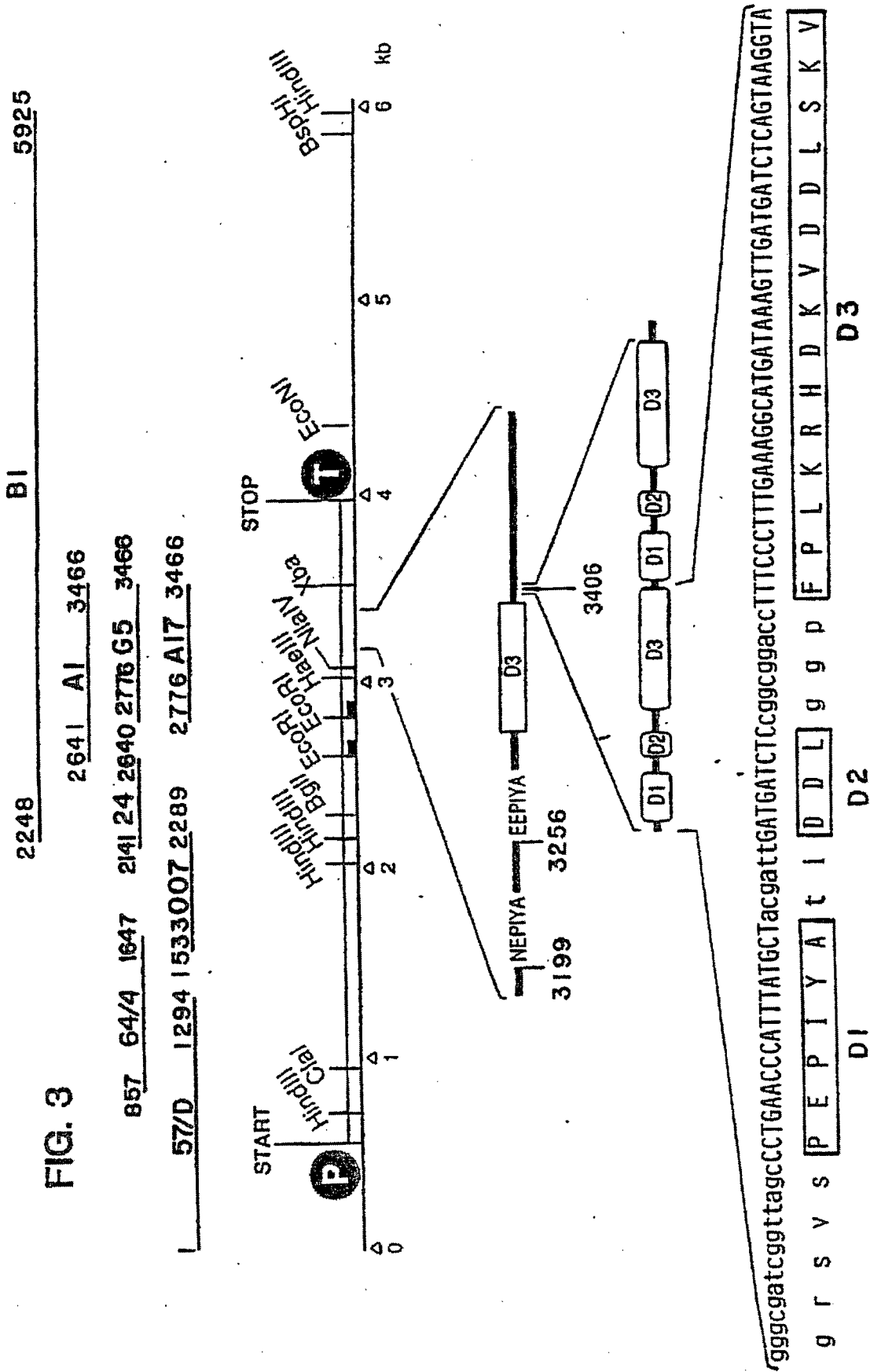
FIG. 1B

2701 TGGCTAACCG CTCTAAAGAT ATTGACACGC TTTATGCTAA CTCAGGCGCT
 2751 CAAGGCAGGG ATCTCTTACA AACCTTATTG ATTGATAGCC ATGATGCGGG
 2801 TTATGCCAGA AAAATGATTG ATGCTACAAG CGCTAATGAA ATCACCAAGC
 2851 AATTGAATAC GGCCACTACC ACTTTAAACA ACATAGCCAG TTTAGAGCAT
 2901 AAAACCAGCG GCTTACAAAC TTTGAGCTTG AGTAATGCGA TGATTTTAAA
 2951 TTCTCGTTTA GTCAATCTCT CCAGGAGACA CACCAACCAT ATTGACTCGT
 3001 TCGCCAAACG CTTACAAGCT TTAAGAGACC AAAAATTCGC TTCTTTAGAA
 3051 AGCGCGGCAG AAGTGTGTA TCAATTTGCC CCTAAATATG AAAAACCTAC
 3101 CAATGTTTGG GCTAACGCTA TTGGGGGAAC GAGCTTGAAT AATGGCTCTA
 3151 ACGCTTCATT GTATGGCACA AGCGCGGGCG TAGACGCTTA CCTTAACGGG
 3201 CAAGTGGAAG CCATTGTGGG CGGTTTTGGA AGCTATGGTT ATAGCTCTTT
 3251 TAATAATCGT GCGAACTCCC TTAAGCTCTG GCGCAATAAC ACTAATTTTG
 3301 GCGTGTATAG CCGTATTTTA ACCAACCAGC ATGAATTTGA CTTTGAAGCT
 3351 CAAGGGGCAC TAGGGAGCGA TCAATCAAGC TTGAATTTCA AAAGCGCTCT
 3401 ATTACAAGAT TTGAATCAA GCTATCATT CTTAGCCTAT AGCGCTGCAA
 3451 CAAGAGCGAG CTATGGTTAT GACTTCGCGT TTTTLAGGAA CGCTTTAGTG
 3501 TTAAAACCAA GCGTGGGTGT GAGCTATAAC CATTTAGGTT CAACCAACTT
 3551 TAAAAGCAAC AGCACCAATC AAGTGGCTTT GAAAAATGGC TCTAGCAGTC
 3601 AGCATTTATT CAACGCTAGC GCTAATGTGG AAGCGCGCTA TTATTATGGG
 3651 GACACTTCAT ACTTCTACAT GAATGCTGGA GTTTTACAAG AGTTCGCTCA
 3701 TGTTGGCTCT AATAACGCCG CGTCTTTAAA CACCTTTAAA GTGAATGCCG
 3751 CTCGCAACCC TTAAATACC CATGCCAGAG TGATGATGGG TGGGGAATTA
 3801 AAATTAGCTA AAGAAGTGTT TTTGAATTTG GCGTTGTTT ATTTGCACAA
 3851 TTTGATTTCC AATATAGGCC ATTTGCTTC CAATTTAGGA ATGAGGTATA
 3901 GTTTCTAAAT ACCGCTCTTA AACCCATGCT CAAAGCATGG GTTTGAAATC
 3951 TTACAAAACA

FIG. 1C

1 MEIQQTHRKI NRPLVSLALV GALVSITPQQ SHAAFFTTVI IPAIVGGIAT
 51 GTAVGTVSGL LSWGLKQAE E ANKTPDKPDK VWRIQAGKGF NEFPNKEYDL
 101 YRSLLSSKID GGWDWGNAAR HYWVKGGQON KLEVDMKDAV GTYTLISGLRN
 151 FTGGDLVDNM QKATLRLGQF NGNSFTSYKD SADRTTRVIS TLKISQLIIL
 201 QKSTTRVGS G AGRKASSTVL TLQASEGITS DKNAEISLYD GATLNLASSS
 251 VKLMGNVWVG RLQYVGAYLA PSYSTINTSK VTGEVNFNHL TVGDKNAAQA
 301 GIIANKKTN I GTLDLWQSAG LNIIAPPEGG YKDKPNNTPS QSGAKNDKNE
 351 SAKNDKQESS QNNSNTQVIN PPNSAQKTEV OPTQVIDGPF AGGKDTVVNI
 401 NRINTNADGT IRVGGFKASL TTNAHLHIG KGGVNLSNQA SGRSLIVENL
 451 TGNITVDGPL RVNNQVGGYA LAGSSANFEE KAGTDTKNGT ATFNNDISLG
 501 RFVNLKVD A TANFKGIDTG NGGFNTLDFS GVTDKVNINK LITASTNVAV
 551 KNFNINELIV KTNGISVGEY THFSEDIGSQ SRINTVRLET GTRSLFSGGV
 601 KFKGGKELVI DEFYYS PWN Y FDARNIKNVE ITNKLAFGPQ GSPWGTSKLM
 651 FNNLT LGQNA VMDYSQFLNL TIQ6DFINQ GTINYLV RGG KVATLSVGNA
 701 AAMMFNNDID SATGFYKPLI KINSAQDLIK NTEHVLLKAK IIGYGNVSTG
 751 TNGISNVNLE EQFKERLALY NNNNRMDTCV VRNTDDIKAC GMAIGDQSMV
 801 NNP DNYKYLI GKAWKNIGIS KTANGSKISV YYLGNSTPTE NGGNTTNLPT
 851 NTTSNARSAN NALAQNA PFA QPSATPNLVA INQHDFGTIE SVFELANRSK
 901 DIDTLYANS G AQGRDLLQTL LIDSHDAGYA RKMIDATSAN EITKQLNTAT
 951 TTLNNIASLE HKTSG LQTLS LSNAMILNSR LVNLSRRHTN HIDSFAKRLQ
 1001 ALKDQKFASL ESAAEVLYQF APKYEKPTNV WANAIGGTSL NNGSNASLYG
 1051 TSAGV DAYLN GQVEAIVGGF GSYGYSSFNN RANSLNSGAN NTNFGVYSRI
 1101 LTNQHEFD FE AQGALGSDQS SLNFKSALLQ DLNQSYHYLA YSAATRASYG
 1151 YDFAFFRNAL VLKPSVGVS Y NHLGSTNFKS NSTNQVALKN GSSSQHLFNA
 1201 SANVEARYYY GDTSYFYMNA GVLQEFHVG SNNAASLNTF KVNAARNPLN
 1251 THARVMMGGE LKLAKEVFLN LGVVYLHNLI SNIGHFASN L GMRYSF

FIG. 2



CTCCATTTTAAGCAACTCCATAGACCACTAAAGAACTTTTTTTGAGGCTATCTTTGAAA
GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACCTATTTTTAGAACGCTT
TCAAAAAGATTCAATTTCTTATTTCTTGTTCTTATTAAGTTCTTTCATTTTAGCAAATTT
CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAAATGCTTGATATTGTTGTAT
TTGACACTAACAAGATACCGATAGGTATGAACTAGGTATAGTAAGGAGAAACAATGACT
M T
AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT
23 N N L Q V A F L K V D N A V A S Y D P D
CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTAT
63 Q L R E E Y S N K A I K N P T K K N Q Y
GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT
103 E S S T K S F Q K F G D Q R Y R I F T S
GAAAATATCATACAACCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT
143 E N I I Q P P I L D D K E K A E F L K S
ATGGGCGTGTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAATGGAGAGCCT
183 M G V F D E S L K E R Q E A E K N G E P
GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCATGTCCAACCAGATATAGCCACT
223 D V K E A I N Q E P V P H V Q P D I A T
AATTTTTCTAAATTCACCTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC
263 N F S K F T L G D M E M L D V E G V A D
TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT
303 L M G S H N G I E P E K V S L L Y G G N
AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA
343 N N V A T I I N V H M K N G S G L V I A
GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTTCATGGAATTT
383 G S Q R A L S Q E E I Q N K I D F M E F
ACTGAGATTAAAGATTTCCAAAAGACTCTAAGGCTTATTTAGACGCCCTAGGGAATGAT
423 T E I K D F Q K D S K A Y L D A L G N D
AATGGGGATTTGAGCTACACTCTCAAAGATTATGGGAAAAAGCAGATAAAGCTTTAGAT
463 N G D L S Y T L K D Y G K K A D K A L D
TATTCTAATTTCAAATACACCAACGCCTCCAAGAATCCCAATAAGGGTGTAGGCGTTACG

FIG. 4A

ATCTGTCCTATTGATTTGTTTTCCATTTTGTTC	120
CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTAGCGTTGCATTTGATTTCT	240
TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG	360
AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTATAGCAGTCT	480
AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATC	600
N E T I D Q Q P Q T E A A F N P Q Q F I	
CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG	720
Q K P I V D K N D R D N R Q A F E G I S	
TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA	840
F S D F I N K S N D L I N K D N L I D V	
TGGGTGTCCCATCAAAACGATCCGTCTAAATCAACACCCGATCGATCCGAAATTTTATG	960
W V S H Q N D P S K I N T R S I R N F M	
GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC	1080
A K Q S F A G I I I G N Q I R T D Q K F	
ACTGGTGGGGATTGGTTGGATATTTTCTCTCATTTATATTGACAAAAACAATCTTCT	1200
T G G D W L D I F L S F I F D K K Q S S	
ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT	1320
T T T D I Q G L P P E A R D L L D E R G	
ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG	1440
I D P N Y K F N Q L L I H N N A L S S V	
GGTGGTCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC	1560
G G P G A R H D W N A T V G Y K D Q Q G	
GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA	1680
G G E K G I N N P S F Y L Y K E D Q L T	
CTTGACAAAAATAATGCTAAATTAGACAACCTTGAGCGAGAAAGAGAAGGAAAAATTCCGA	1800
L A Q N N A K L D N L S E K E K E K F R	
CGTATTGCTTTTGTTCCTAAAAAAGACACAAAACATTCAGCTTTAATTACTGAGTTTGGT	1920
R I A F V S K K D T K H S A L I T E F G	
AGGGAGAAAAATGTTACTCTTCAAGGTAGCCTAAACATGATGGCGTGATGTTTGTGAT	2040
R E K N V T L Q G S L K H D G V M F V D	
AATGGCGTTTCCCATTTAGAAGTAGGCTTTAACAAGGTAGCTATCTTTAATTTGCCTGAT	2160

FIG. 4B

503 Y S N F K Y T N A S K N P N K G V G V T
 TTAAATAATCTCGCTATCACTAGTTTCGTAAGGCGGAATTTAGAGGATAAACTAACCCT
 543 L N N L A I T S F V R R N L E D K L T T
 GAATTGGTTGGAAAACTTTAACTTCAATAAAGCTGTAGCTGACGCTAAAAACACAGGC
 583 E L V G K T L N F N K A V A D A K N T G
 CATTTAGAGAAAGAAGTAGAGAAAAAATTGGAGAGCAAAAGCGGCAACAAAAATAAAATG
 623 H L E K E V E K K L E S K S G N K N K M
 GCTAATAGAGACGCAAGAGCAATCGCTTACGCTCAGAATCTTAAAGGCATCAAAAGGGAA
 663 A N R D A R A I A Y A Q N L K G I K R E
 GAATTCAAAAATGGCAAAAATAAGGATTTTCAGCAAGGCAGAAGAACTAAAGCCCTT
 703 E F K N G K N K D F S K A E E T L K A L
 AATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTTCAGCAAGGTAACGCAA
 743 N A A L N E F K N G K N K D F S K V T Q
 AAAGTTGATAATCTCAATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTTCAGTAGG
 783 K V D N L N Q A V S V A K A T G D F S R
 CAAAAAATGAAAGTCTCAATGCTAGAAAAAATCTGAAATATATCAATCCGTAAAGAAT
 823 Q K N E S L N A R K K S E I Y Q S V K N
 AAAAATTTTCGGACATCAAGAAAGAGTTGAATGCAAACTTGGAAATTTCAATAACAAT
 863 K N F S D I K K E L N A K L G N F N N N
 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA
 903 Q A A S L E E P I Y A Q V A K K V N A K
 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA
 943 P L K R H D K V D D L S K V G L S R N Q
 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG
 983 F G N L E Q T I D K L K D S T K H N P M
 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA

FIG.4C

N G V S H L E V G F N K V A I F N L P D
 AAAGGATTGTCCCCACAAGAAGCTAATAAGCTTATCAAAGATTTTTTGAGCAGCAACAAA 2280
 K G L S P Q E A N K L I K D F L S S N K
 AATTATGATGAAGTGAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAG 2400
 N Y D E V K K A Q K D L E K S L R K R E
 GAAGCAAAAGCTCAAGCTAACAGCCAAAAAGATGAGATTTTTGCGTTGATCAATAAAGAG 2520
 E A K A Q A N S Q K D E I F A L I N K E
 TTGTCTGATAAACTTGAAAATGTCAACAAGAATTTGAAAGACTTTGATAAATCTTTTGAT 2640
 L S D K L E N V N K N L K D F D K S F D
 AAAGGTTTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTT 2760
 K G S V K D L G I N P E W I S K V E N L
 GCAAAAAGCGACCTTGAAAATTCGGTTAAAGATGTGATCATCAATCAAAAGGTAACGGAT 2880
 A K S D L E N S V K D V I I N Q K V T D
 GTAGAGCAAGCGTTAGCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCCAACAAGCT 3000
 V E Q A L A D L K N F S K E Q L A Q Q A
 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACCTCTTTCT 3120
 G V N G T L V G N G L S Q A E A T T L S
 AACATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 3240
N N N G L K N E P I Y A K V N K K K A G
 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 3360
 I D R L N Q I A S G L G V V G Q A A G F
 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 3480
 E L A Q K I D N L N Q A V S E A K A G F
 AATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTGAGCGAAACTAGACAAT 3600
 N L W V E S A K K V P A S L S A K L D N
 AAAGCGACCGGCATGCTAACGCAAAAAAACCTGAGTGGCTCAAGCTCGTGAATGATAAG 3720

FIG.4D

1023 Y A T N S H I R I N S N I K N G A I N
 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTC
 1063 I V A H N V G S V P L S E Y D K I G F
 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTAAACCAATGCATTTTCTACAGCA
 1103 V K D T N S G F T Q F L T N A F S T A
 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCCACCCCTTG
 1143 G F Q K S
 TGAATGCTACCAATTCATGGTATCATATCCCATACATTCGTATCTAGCGTAGGAAG
 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGG
 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCTACCCC
 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAACTAGAAAAAAGAAGTAGCTATA
 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAGACGCT
 ATGCCTTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACCTTG
 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTATGAAGCGATGAGAAGA
 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACAC
 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTCTAGCCAGATCACATC
 CTGCAAATATCCTACAATAGCATCGCCGAATGGATGAGTAGGGGGGGTGTGAAAG
 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTT
 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGG
 GCTAAAATGCTTGGCTCAATCACGCCACAATAGGGATTTTGAATGCTTTTGCATC
 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTG
 AGGCTTTTAAACGCTAAACCCTCCACACCGCTATCAAAAACGCCTATTTTCATG
 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 5925

FIG. 4E

E K A T G M L T Q K N P E W L K L V N D K
 AACCAGAAGAATATGAAAGATTATTCTGATTTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 3840
 N Q K N M K D Y S D S F K F S T K L N N A
 TCTTATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 3960
 S Y Y C L A R E N A E H G I K N V N T K G
CTAAAAGCGAGGGGTTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 4080

 TGTGCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 4200
 GCAACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 4320
 AAAAAAGACTTAACCCTTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTTCGTGCTATAGTG 4440
 ACAAAGATCAAGTTCAAAAATCATAGAGCTTTTAGAGCAAATTGATCGCGCTCTTAACCAA 4560
 TGCGATCAGAAGTGGA AAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACTGTGG 4680
 GTAATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 4800
 ATATCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 4920
 AATACCCTTATAGTGTGAGCTATAGCCCCCTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 5040
 GCCGCTCGCATGAAATTCACCTTTAGGGAATGCGTGTGCATTTTTTTTAAGGGCGTATTTTTG 5160
 GGCAAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 5280
 TCTAACAAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAGAGAAGTGGCTAAA 5400
 ATCGTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTTACTTGTCGCTTGATC 5520
 TCTTCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 5640
 CCATAAGGCACTCTAGCCGTATCGCCATAATAGATGATTTTCATCAAATAATTGCGCTTTTAAA 5760
 ACACTTTTTTAATTTAATGGGATTAATTAAGGATTTTATTTTTTCATTCATTAAGTTTAAAAAT 5880

FIG. 4 F

10 30 50
 AAGCTTGCTGTCATGATCACAAAAACACTAAAAACATTATTATTAAAGGATACAAAATG
 M
 70 90 110
 GCAAAAGAAATCAAATTTTCAGATAGTGCGAGAAACCTTTTATTTGAAGGCGTGAGGCAA
 A K E I K F S D S A R N L L F E G V R Q
 130 150 170
 CTCCATGACGCTGTCAAAGTAACCATGGGGCCAAGAGGCAGGAATGTATTGATCCAAAAA
 L H D A V K V T M G P R G R N V L I Q K
 190 210 230
 AGCTATG6CGCTCCAAGCATCACCAAGACGGCGTGAGCGTGGCTAAAGAGATTGAATTA
 S Y G A P S I T K D G V S V A K E I E L
 250 270 290
 AGTTGCCCAGTAGCTAACATGGGCGCTCAACTCGTTAAAGAAGTAGCGAGCAAAACCGCT
 S C P V A N M G A Q L V K E V A S K T A
 310 330 350
 GATGCTGCCGGCGATGGCAGCACCACAGCGACCGTGCTAGCTTATAGCATTTTTAAAGAA
 D A A G D G T T T A T V L A Y S I F K E
 370 390 410
 GGTTTGAGGAATATCACGGCTGGGGCTAACCCTATTGAAGTGAAACGAGGCATGGATAAA
 G L R N I T A G A N P I E V K R G -M D K
 430 450 470
 GCTGCTGAAGCGATCATTAAATGAGCTTAAAAAAGCGAGCAAAAAAGTAGGCGGTAAAGAA
 A A E A I I N E L K K A S K K V G G K E
 490 510 530
 GAAATCACCCAAGTGGCGACCATTCTGCAAACTCCGATCACAATATCGGGAAACTCATC
 E I T Q V A T I S A N S D H N I G K L I
 550 570 590
 GCTGACGCTATGGAAAAAGTGGGTAAAGACGGCGTGATCACCGTTGAGGAAGCTAAGGGC
 A D A M E K V G K D G V I T V E E A K G
 610 630 650
 ATTGAAGATGAATTGGATGTCGTAGAAGGCATGCAATTTGATAGAGGCTACCTCTCCCT
 I E D E L D V V E G M Q F D R G Y L S P

FIG. 5A

```

        670                690                710
TATTTTGTAAACGACGCTGAGAAAATGACCGCTCAATTGGATAATGCTTACATCCTTTTA
Y F V T N A E K M T A Q L D N A Y I L L
        730                750                770
ACGGATAAAAAAATCTCTAGCATGAAAGACATTCTCCCGCTACTAGAAAAAACCATGAAA
T D K K I S S M K D I L P L L E K T M K

        790                810                HindIII
GAGGGCAAACCGCTTTTAAATCATCGCTGAAGACATTGAGGGCGAAGCTTTAACGACTCTA
E G K P L L I I A E D I E G E A L T T L
        850                870                890
GTGGTGAATAAATTAAGAGGCGTGTTGAATATCGCAGCGGTTAAAGCTCCAGGCTTTGGG
V V N K L R G V L N I A A V K A P G F G
        910                930                950
GACAGAAGAAAAGAAATGCTCAAAGACATCGCTATTTTAAACCGGCGGTCAAGTCATTAGC
D R R K E M L K D I A I L T G G Q V I S
        970                990                1010
GAAGAATTGGGCTTGAGTCTAGAAAACGCTGAAGTGGAGTTTTTAGGCAAAGCTGGAAGG
E E L G L S L E N A E V E F L G K A G R
        1030                1050                1070
ATTGTGATTGACAAAGACAACACCACGATCGTAGATGGCAAAGGCCATAGCGATGATGTT
I V I D K D N T T I V D G K G H S D D V
        1090                1110                1130
AAAGACAGAGTCGCGCAGATCAAAACCCAAATTGCAAGTACGACAAGCGATTATGACAAA
K D R V A Q I K T Q I A S T T S D Y D K
        1150                1170                1190
GAAAAATTGCAAGAAAGATTGGCTAAACTCTCTGGCGGTGTGGCTGTGATTAAAGTGGGC
E K L Q E R L A K L S G G V A V I K V G
        1210                1230                1250
GCTGCGAGTGAAGTGGAAATGAAAGAGAAAAAAGACCGGGTGGATGACGCGTTGAGCGCG
A A S E V E M K E K K D R V D D A L S A
        1270                1290                1310
ACTAAAGCGGCGGTTGAAGAAGGCATTGTGATTGGTGGCGGTGCGGCTCTCATTCGCGCG
T K A A V E E G I V I G G G A A L I R A

```

FIG. 5B

```

      1330              1350              1370
GCTCAAAAAGTGCATTTGAATTTGCACGATGATGAAAAAGTGGGCTATGAAATCATCATG
A Q K V H L N L H D D E K V G Y E I I M
      1390              1410              1430
CGCGCCATTAAAGCCCCATTAGCTCAAATCGCTATCAACGCTGGTTATGATGGCGGTGTG
R A I K A P L A Q I A I N A G Y D G G V
      1450              1470              1490
GTCGTGAATGAAGTAGAAAAACACGAAGGGCATTTTGGTTTTAACGCTAGCAATGGCAAG
V V N E V E K H E G H F G F N A S N G K
      1510              1530              1550
TATGTGGATATGTTTAAAGAAGGCATTATTGACCCCTTAAAAGTAGAAAGGATCGCTCTA
Y V D M F K E G I I D P L K V E R I A L
      1570              1590              1610
CAAATGCGGTTTTCGGTTTCAAGCCTGCTTTTAACACAGAAGCCACCGTGCATGAAATC
Q N A V S V S S L L L T T E A T V H E I
      1630              1650              1670
AAAGAAGAAAAAGCGACTCCGGCAATGCCTGATATGGGTGGCATGGGCGGTATGGGAGGC
K E E K A T P A M P D M G G M G G M G G
      1690              1710              1730
ATGGGCGGCATGATGTAAGCCCGCTTGCTTTTGTAGTATAATCTGCTTTTAAATCCCTTC
M G G M M *
      1750              1770              1790
TCTAAATCCCCCCTTTCTAAAATCTCTTTTTTGGGGGGGTGCTTTGATAAACCGCTCG

      1810              1830
CTTGTAACAAACATGCAACAAAAATCTCTGTTAAGCTT

```

FIG. 5C