



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) **DE 693 33 992 T2** 2006.11.30

(12)

## Übersetzung der europäischen Patentschrift

(97) **EP 0 728 015 B1**

(21) Deutsches Aktenzeichen: **693 33 992.6**

(86) PCT-Aktenzeichen: **PCT/US93/05899**

(96) Europäisches Aktenzeichen: **93 916 641.9**

(87) PCT-Veröffentlichungs-Nr.: **WO 1994/000151**

(86) PCT-Anmeldetag: **18.06.1993**

(87) Veröffentlichungstag

der PCT-Anmeldung: **06.01.1994**

(97) Erstveröffentlichung durch das EPA: **28.08.1996**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **08.03.2006**

(47) Veröffentlichungstag im Patentblatt: **30.11.2006**

(51) Int Cl.<sup>8</sup>: **C07K 14/16** (2006.01)

**C07K 1/04** (2006.01)

**C07K 17/00** (2006.01)

**C12N 15/00** (2006.01)

(30) Unionspriorität:

**900123**      **18.06.1992**      **US**

(73) Patentinhaber:

**Creagen, Inc., Cambridge, Mass., US**

(74) Vertreter:

**Luderschmidt, Schüler & Partner, 65189  
Wiesbaden**

(84) Benannte Vertragsstaaten:

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

(72) Erfinder:

**CREA, Roberto, Belmont, MA 02178, US**

(54) Bezeichnung: **VERFAHREN ZUR HERSTELLUNG EINES SETS VON KOMBINATORISCHEN POLYPEPTIDANTI-  
GENEN**

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**

**[0001]** Der Schutz eines Wirts ist ein fundamentales Kennzeichen des Immunsystems von Vertebraten. Zu diesem Zweck erfüllen Antikörper zahlreiche Funktionen bei der Abwehr von Krankheitserregern. Antikörper können beispielsweise ein biologisch aktives Molekül neutralisieren, den Stoffwechselweg des Komplementsystems induzieren, eine Phagozytose stimulieren (Opsonisierung) oder an der antikörperabhängigen zellvermittelten Cytotoxizität (ADCC) teilnehmen.

**[0002]** Falls sich der Antikörper an einen Ort bindet, der für die biologische Funktion eines Moleküls kritisch ist, kann die Aktivität des Moleküls neutralisiert werden. Auf diese Weise können spezifische Antikörper die Bindung eines Virus oder eines Protozoons an die Oberfläche einer Zelle blockieren. Auf ähnliche Weise können Bakterien und andere Arten von Toxinen von geeigneten Antikörpern gebunden und neutralisiert werden. Unabhängig davon, ob ein gebundener Antikörper sein Ziel neutralisiert oder nicht, kann der erhaltene Antigen-Antikörper-Komplex darüber hinaus mit anderen Abwehrmechanismen in Wechselwirkung treten, was zu einer Zerstörung und/oder Beseitigung des Antigens führt.

**[0003]** Parasiten haben ein breites Spektrum von Mechanismen entwickelt, um eine Immunantwort zu vermeiden. Eine antigene Variation ist vielleicht eine der am meisten untersuchten Vermeidungsstrategien, zum Teil deshalb, weil eine solche Variation die Entwicklung eines Impfstoffs besonders schwierig macht. Im Allgemeinen gibt es zwei Wege, auf denen eine antigene Variation auftreten kann: eine Antigendrift und ein Antigenwechsel. Eine Antigendrift ist relativ einfach. Punktmutationen führen zu Genen, welche pathogene Antigene codieren, indem sie einige der Epitope auf dem Antigen verändern, so dass das immunologische Gedächtnis des Wirts für das ursprüngliche Antigen durch die Mutante nicht mehr ausgelöst wird. Da eine Immunität gegen eine Variante nicht notwendigerweise eine Immunität gegen andere Varianten garantiert, kann eine Häufung solcher Punktmutationen in einer pathogenen Population im gleichen Wirt zu Mehrfachinfektionen führen. Eine Antigendrift ist bei den meisten Pathogenen gefunden worden (einschließlich bei Viren, Bakterien und Protozoen), wobei ihre Bedeutung bei den einzelnen Spezies variiert.

**[0004]** Viele Viren sind zu großen antigenen Variationen in der Lage und es sind eine große Zahl von serologisch unterschiedlichen Stämmen dieser Viren identifiziert worden. Dies führt dazu, dass ein besonderer Virusstamm gegen die in der Population durch eine vorherige Infektion oder Impfung erzeugte Immunität unempfindlich wird. So wurde z.B. der Fortschritt bei der Entwicklung eines HIV-1-Impfstoffs durch die Veränderbarkeit der Aminosäuresequenz bei verschiedenen isolierten HIV-1-Stämmen behindert. Diese Variabilität ist bei der äußeren Proteinhülle gp 120 besonders groß, welche das primäre Ziel für Antikörper ist, welche die Ansteckungsfähigkeit durch ein Virus neutralisieren (Robey et al. (1986) PNAS 83:7023; Putney et al. (1986) Science 234:1392 und Rusche et al. (1987) PNAS 84:6924). Untersuchungen am Menschen und der Maus haben einen kleinen Abschnitt des gp 120 ergeben, der als V3-Loop oder Principal Neutralizing Determinant (PND) bezeichnet wird und etwa 35 Reste zwischen zwei Invarianten über eine Disulfid-Quervernetzung verbundene Cysteine (Cys-303 bis Cys-338: HIV-1-Nomenklatur von Takahashi et al. (1992) Science 255: 333) umfasst, welcher die hauptsächlich neutralisierenden Antikörper zum Virus lenkt (Palker et al. (1988) PNAS 85:1932; Rusche et al. (1988) PNAS 85:3198 und Goudsmit et al. (1988) PNAS 85:4478). Während derselbe Abschnitt in unterschiedlichen Klonisolaten hinsichtlich seiner Sequenz die meiste Variabilität aufweist (Takahashi et al. (1992) Science 255:333) ergab eine Analyse der Aminosäuresequenzen dieser Domäne in 8 von 14 Positionen im zentralen Abschnitt des V3-Loops eine Konservierung von über 80% der Aminosäuren, was nahe legt, dass es für die Variabilität des V3-Loops Beschränkungen gibt (LaRosa et al. (1990) Science 249:932). Wegen dieser Variabilität neutralisieren die von der PND aus einem Isolat induzierten neutralisierenden Antikörper keine Isolate mit PNDs von unterschiedlicher Aminosäuresequenz.

**[0005]** Auch Versuche, Influenza durch Impfung unter Kontrolle zu bringen, waren bisher von begrenztem Erfolg und wurden durch die kontinuierlichen Änderungen bei den hauptsächlich Oberflächenantigenen der Influenzaviren, dem Hämagglutinin (HA) und der Neuraminidase (NA) behindert, gegen welche sich die neutralisierenden Antikörper primär richten (Caton et al. (1982) Cell 31:417; Cox et al. (1983) Bulletin WHO 61:143; Eckert, E.A. (1973) J. Virology 11:183. Die Influenzaviren verfügen über die Fähigkeit, innerhalb einer kurzen Zeit ihre Antigene in hohem Grade zu variieren. Es ist diese Eigenschaft des Virus, die es so schwierig macht, die Kontrolle über das jahreszeitliche Auftreten von Influenza bei Populationen von Mensch und Tier zu gewinnen.

**[0006]** Mit Hilfe serologischer und sequenzierender Untersuchungen konnten beim Influenza A-Virus zwei Arten von antigenen Variationen nachgewiesen werden. Eine Antigenshift tritt primär auf, wenn in einem neuen Virus-Stamm entweder HA oder NA oder beide durch ein antigenetisch neuartiges HA oder NA ersetzt werden.

Das von einer Antigenshift hervorgerufene Vorkommen neuer Subtypen führt gewöhnlich zu pandemischen Infektionen.

**[0007]** Eine Antigendrift tritt bei Influenzaviren eines gegebenen Subtyps auf. Sequenzanalysen von Aminosäuren und Nucleotiden legen nahe, dass eine Antigendrift durch eine Reihe aufeinander folgender Mutationen auftritt, was zu Änderungen der Aminosäuren im Polypeptid und zu Unterschieden in den antigenen Eigenschaften des Virus führt. Die Anhäufung verschiedener Mutationen mittels Antigendrift ergibt ggf. einen Subtyp, der in der Lage ist, die Immunantwort einer großen Zahl von Subjekten zu entgehen, die zuvor einem ähnlichen Subtyp ausgesetzt gewesen waren. Tatsächlich sind ähnliche neue Varianten durch die Passage von Viren in Mäusen oder Hühnerembryonen in Gegenwart geringer Mengen von Antikörpern selektiert worden. Eine Antigendrift führt zu weniger ernstesten Ausbrüchen von Epidemien oder Infektionen. Eine Antigendrift ist auch bei Influenza B-Viren beobachtet worden.

**[0008]** Der derzeit in Klinikversuchen verwendete, gegen das Hepatitis B-Virus gerichtete gereinigte Antigenimpfstoff besteht aus Antigenen eines einzelnen viralen Subtyps. Der Entscheidung dafür lag die Überlegung zu Grunde, dass sowohl für die S-Region als auch für die Prä-S(2)-Region spezifische Antikörper, von denen gezeigt worden war, dass sie bei der Neutralisierung des Hepatitis B-Virus eine geringe Wirksamkeit zeigten, in erster Linie gruppenspezifisch sind. In dieser Betrachtungsweise wird jedoch der Einfluss des viralen Subtyps auf die T-Zell-Erkennung nicht berücksichtigt. Es ist gezeigt worden, dass die Prä-S(2)-spezifische T-Zell-Antwort bei der Maus für den Subtyp hoch spezifisch ist (Milic H. et al. (1990) J.Immunol. 144: 3535).

**[0009]** Das einzellige Protozoon Plasmodium falciparum ist beim Menschen das vorherrschende Pathogen für die Verursachung der Malaria. Die Infektion beginnt, wenn in der Speicheldrüse von Anopheles-Moskitos vorkommende Sporoziten in das Blut von Wirtsempfängern gelangt. Die Sporoziten dringen rasch in die Hepatocyten ein, in denen sie sich weiter zu Schizonten entwickeln. Nach der Reifung werden infektiöse Merozoiten in das Blut des Wirts freigesetzt und dringen in die Erythrozyten ein und beginnen einen neuen schizogamen Vermehrungszyklus, welcher mit den klinischen Symptomen der Malaria einhergeht. Die Zahl der von der Weltgesundheitsorganisation vorausgesagten Malariafälle liegt weltweit bei über 100 Millionen.

**[0010]** Über den Schutz von Affen gegen eine Infektion durch bestimmte Arten von Plasmodium durch Immunisierung mit gereinigten Oberflächenantigenen, die während mindestens einer Stufe des Lebenszyklus des Parasiten exprimiert worden waren, wurde nur von einem beschränkten Erfolg berichtet. Beispielsweise stellt das als p190 oder polymorphes Schizontenantigen bezeichnete Merozoitenprotein einen vielversprechenden Kandidaten für einen in das Blut zu verabreichenden Impfstoff dar (Herra et al. (1992) Infection and Immunity 60:154–158); Merkli et al. (1984) Nature 311:379–382; Mackay et al. (1985) EMBO J, 4:3823–3829). Das Antigen p190 ist ein großes Glykoprotein, das während der Bildung der Merozoiten synthetisiert und extensiv weiter umgebaut wird, wobei das Umbauprodukt von 80.000 Da das Haupthüllprotein der reifen Merozoiten ist. Sonden aus monoklonalen Antikörpern gegen p190 und die Analyse der Primärsequenz zeigen, dass das Antigen polymorphe Sequenzen enthält, welche bei den Spezies und Subspezies von Plasmodium eine antigene Variation hervorrufen.

#### Zusammenfassung der Erfindung

**[0011]** Diese Erfindung betrifft einen Satz von Polypeptid-Antigenen mit Aminosäuresequenzen, die sich von Aminosäuresequenzen aus einer Population von Varianten eines Proteins oder eines Teils davon ableiten, sowie Verfahren zur Herstellung des Satzes von Polypeptid-Antigenen. Im Allgemeinen umfasst das Verfahren

- a. die Auswahl eines Proteins oder eines Teils davon, welche eine Population von N Varianten aufweisen, dargestellt durch die Formel  $A_2 A_3 \dots A_{n-2} A_{n-1} A_n$ , wobei  $A_n$  eine Aminosäure ist, die an der Aminosäure-Position n des Proteins oder des Teils davon vorkommt;
- b. die Bestimmung der Anzahl der Male  $O_n^{aa}$ , die jeder Typ von Aminosäure an jeder Aminosäure-Position n in den Z Varianten vorkommt;
- c. die Berechnung der Häufigkeit des Vorkommens  $(O_n^{aa}/N)_n$  jedes Typs von Aminosäure an jeder Aminosäure-Position n in den N Varianten; und
- d. die Synthese eines Satzes von Polypeptid-Antigenen mit den im Wesentlichen durch die Formel  $A'_1 A'_2 A'_3 \dots A'_{n-2} A'_{n-1} A'_n$  wiedergegebenen Aminosäure-Sequenzen, wobei  $A'_n$  als ein Typ von Aminosäure definiert ist, der in größerer als einer ausgewählten Häufigkeit an der entsprechenden Aminosäure-Position in den N Varianten vorkommt.

**[0012]** In einer bevorzugten Ausführungsform wird der Satz von Polypeptid-Antigenen erzeugt, indem eine degenerierte Oligonucleotid-Sequenz mit einer Mindestanzahl von Nucleotid-Kombinationen an jeder Co-

don-Position  $n$  ermittelt wird, welche Kombinationen wenigstens die Codons umfassen, die für jeden Typ von Aminosäure  $A'_1$  bis  $A'_n$  codieren. Das degenerierte Oligonucleotid wird in ein exprimierbares Gen eingebaut, um einen Satz von Genen zu erzeugen. Der Satz von Genen wird in einem geeigneten Expressionssystem exprimiert, um den Satz von Polypeptid-Antigenen

$$A'_1 A'_2 A'_3 \dots A'_{n-2} A'_{n-1} A'_n$$

zu erzeugen.

**[0013]** Typischerweise reicht  $n$  von 8 bis etwa 150. Die für einen Einbau einer Aminosäure in den Satz von Polypeptid-Antigenen an einer vorgegebenen Position ausgewählte Grenze für die Häufigkeit kann für alle Aminosäure-Positionen auf den gleichen Wert festgelegt werden. Alternativ kann die ausgewählte Grenze für die Häufigkeit für jede Aminosäure-Position einzeln festgelegt und von Position zu Position geändert werden. Typischerweise reicht eine Grenzhäufigkeit von 5%–15%. Dies bedeutet, dass für den Einbau einer bestimmten Aminosäure an einer vorgegebenen Position in dem Satz der erzeugten Polypeptid-Antigene die Aminosäure in dieser Position mit mehr als 5%–15% der N-Varianten erscheinen muss (d.h.  $O_n^{aa}/N$  muss größer als 5%–15% sein).

**[0014]** Der Satz von Polypeptid-Antigenen kann einer vollständigen Peptid-Sequenz eines Proteins oder nur einem Teil derselben entsprechen. Zusätzlich braucht die variierende Sequenz nicht durchgängig im Protein vorzukommen; sie kann innerhalb eines einzelnen Proteins oder einer Proteinuntereinheit verstreut oder in mehr als einem Protein oder einer Proteinuntereinheit vorliegen.

**[0015]** Der Satz von Polypeptiden kann als künstlicher Impfstoff eingesetzt werden. Die Antigene können dem Wirtsorganismus in einem physiologisch verträglichen Vehikel und in einer Dosierungsabfolge verabreicht werden, die genügt, um einen Immunschutz gegen eine Population von Antigen-Varianten aufzubauen. Ist z.B. das Protein eine Komponente eines Pathogens, kann das Protein oder ein Teil davon Sequenzen mit einem oder mehreren Epitopen umfassen, so dass der Satz von Polypeptid-Antigenen bei Verabreichung als ein Immunogen die Bereitstellung von neutralisierenden Antikörpern gegen das Pathogen durch einen Wirtsorganismus zur Folge hat.

**[0016]** Alternativ lässt sich sowohl die Form als auch der Weg der Verabreichung des Satzes von Polypeptid-Antigenen so einstellen, dass sie Toleranz-induzierend werden und dadurch in einem Wirtsorganismus eine Toleranz gegen ein variierendes Protein-Antigen oder einen Teil desselben aufbauen.

#### Genauere Beschreibung der Erfindung

**[0017]** Eine von einem Pathogen hervorgerufene entweder klinisch erkennbare oder verborgene Infektion kann zu einer Immunität führen und die Immunität gegen Pathogene der gleichen Antigenstruktur erscheint dann als lang anhaltend. Eine erneute Infektion durch das Pathogen kann jedoch von Varianten mit kleinen antigenen Unterschieden verursacht werden. Da eine Immunität in hohem Maße epitopspezifisch ist, wird eine künstlich induzierte Immunität gegen ein Pathogen oft durch eine ausgeprägte Antigenvariation des Pathogens eingeschränkt. Daher hat die Fähigkeit eines Pathogens, einer Antigendrift zum unterliegen, oft die Unwirksamkeit eines herkömmlichen Impfstoffs zur Folge.

**[0018]** Diese Erfindung stellt ein Verfahren zur Herstellung eines Satzes von Polypeptid-Antigenen zur Verfügung, die von einem Protein (oder einem Teil davon) stammen, das unter natürlich oder künstlich induzierten Varianten des Proteins mit einem gewissen Grad von Sequenz-Heterogenität exprimiert wird. Es ist das Ziel, eine Mischung von Antigenen zur Verfügung zu stellen, welche zur Immunisierung gegen Varianten und vorzugsweise mögliche unbekannte oder neue Varianten, die auftreten können, eingesetzt werden kann.

**[0019]** Nach dem Verfahren dieser Erfindung (Ansprüche 1–11) wird in der Gesamtzahl von Varianten für jede Aminosäure-Position des Proteins (oder eines Teils des Proteins) die Häufigkeit des Auftretens eines jeden Aminosäuretyps ermittelt. Aminosäuren, die in der Gesamtzahl der Varianten an jeder Position oberhalb einer vorbestimmten Häufigkeit (z.B. 5%, 10%, 15% usw.) auftreten, werden zum Einsatz ausgesucht, um den Satz von Polypeptid-Antigenen zu erzeugen.

**[0020]** Im Allgemeinen weisen die Polypeptide eine Länge im Bereich von 8 bis 150 Aminosäuren, vorzugsweise im Bereich von 10 bis 50 Aminosäuren auf.

**[0021]** In der bevorzugten Ausführungsform wird der Satz von Polypeptid-Antigenen mit Hilfe eines degenerierten Oligonucleotids gewonnen. Die Sequenz des degenerierten Oligonucleotids lässt sich ermitteln, so dass für jedes Codon, welches notwendig ist, um jeden für einen Einsatz ausgesuchten Aminosäuretyp zu erzeugen (d.h. solche, die mit einer größeren als der gewählten Häufigkeit in der Gesamtzahl der Variationen an der entsprechenden Position für die Aminosäure auftreten), sich ein Minimum an Nucleotidkombinationen ergibt.

**[0022]** Die Mischung aus synthetischen Oligonucleotiden kann enzymatisch in Gensequenzen ligiert werden, so dass der Satz von Polypeptid-Antigenen als einzelne Polypeptide exprimiert werden kann oder als ein Satz von größeren Fusionsproteinen, die darin den Satz von Polypeptid-Antigenen enthalten. Alternativ lässt sich der Satz von Polypeptid-Antigenen mit Hilfe einer organisch-chemischen Peptidsynthese gewinnen. Jede Runde einer Aminosäurekopplung kann so ausgeführt werden, dass an einer vorgegebenen Aminosäureposition eine bestimmte Heterogenität von Aminosäuretypen erhalten wird (indem in den passenden Schritten in der Synthese mehr als eine aktivierte Aminosäure eingesetzt wird). Die Mischung der Aminosäuren beruht auf der Analyse der Häufigkeit für die N Varianten. Alternativ kann die Mischung der Aminosäuren auf der Grundlage der Nucleotid (Codon)-Kombination ermittelt, werden, die bei der Herstellung des degenerierten Oligonucleotids erzeugt wurde.

**[0023]** Die sich aus der Verwendung von degenerierten Nucleotidsequenzen ergebende kombinatorische Wirkung führt typischerweise zu einigen Aminosäuretypen, die nicht in der ursprünglichen Gesamtzahl von Varianten vorkommen. Dies führt in dem gesamten Satz von Polypeptid-Antigenen zur Gewinnung von Polypeptiden, die in der Natur nicht vorgekommen sind. Weil diese Nucleotid-Kombinationen anfangs auf bekannten Varianten basieren, stellen die für die zusätzlichen Aminosäuren verantwortlichen Codons möglicherweise Mutationen dar, die in der Natur wahrscheinlicher sind als die künstlich mittels Strukturanalyse des Proteins hergestellten. Somit kann der Satz von Polypeptid-Antigenen zu einer Immunität gegen einen großen Bereich von sowohl möglicher Varianten als auch von bekannten Varianten des Proteins führen.

**[0024]** Um die Sequenzen einer Gesamtzahl von Varianten eines Proteins zu analysieren, können die in Frage kommenden Aminosäuresequenzen bezüglich ihrer Sequenzhomologie aufgereiht werden. Das Vorkommen oder Fehlen von Aminosäuren aus einer aufgereihten Sequenz einer bestimmten Variante wird zu einer gewählten Konsensuslänge einer Referenzsequenz, die wirklich oder künstlich sein kann, in Bezug gesetzt. Um in einer Aufreihung von Sequenzen die höchste Homologie beizubehalten, können Deletionen in der Sequenz einer Variante in Bezug zu der Referenzsequenz durch einen Aminosäure-Freiraum (\*) wiedergegeben werden, während Insertions-Maturationen in der Variante in Bezug auf die Referenzsequenz beim Aufreihen außer Acht gelassen und in der Sequenz für die Variante weggelassen werden können. Unten werden z.B. zwei mögliche Aufreihungen von drei Sequenzen für den V3-Loop des HIV-Isolats von bekanntem Tropismus angegeben (Hwang et al. (1991) Science 253: 71-76).

**[0025]** Die Sequenzen

BaL . -CTRPNNNTTRKSIHIGPGRALYTTGEIIGDIRQAHC- (SeqIDNo. 1)

HTLV-III B . -CTRPNNNTRKKIRIQRGPGRAFVTIGKIGNMRQAHC- (Seq ID No. 2)

SF162 -CTRPNNNTRKSITIGPGRAFYATGDIIGDIRQAHC- (Seq ID No. 3)

können aufgereiht werden als:

	1	10	20	30	
BaL	-CTRPNNNTRKSIHI	**G	PGRALYTTGEI	IIGDIRQAHC-	(Seq ID No. 1)
HTLV-III B	-CTRPNNNTRKKIRIQ	RGPGRAFVTIGK	*IGNMRQAHC-		(Seq ID No. 2)
SF162	-CTRPNNNTRKSITI	**G	PGRAFYATGDI	IIGDIRQAHC-	(Seq ID No. 3)

in welcher die Reste 15 und 16 des ursprünglichen HTLV-III B-Stamms in der Aufreihung enthalten sind, oder alternativ als:

	1	10	20	30	

BaL -CTRPNNNTRKSIHIGPGRALYTTGEIIGDIRQAHC- (Seq ID No. 1)  
 HTLV-III B -CTRPNNNTRKKIRIGPGRAFVTIGK\*IGNMRQAHC- (Seq ID No. 4)  
 SF162 -CTRPNNNTRKSITIGPGRAFYATGDIIGDIRQAHC- (Seq ID No. 3)

in welcher die Reste 15 und 16 des ursprünglichen HTLV-III B-Stamms aus der Aufreihung der Sequenzen entfernt wurden.

**[0026]** Bei vorgegebenen N Varianten für das Protein und der Anzahl der Male  $O_n^{aa}$ , mit welchen eine gegebene Aminosäure (aa) an einer gegebenen Position n vorkommt, berechnet sich die Häufigkeit des Vorkommens dieser Aminosäure an dieser Position zu  $O_n^{aa}/N$ . Die Häufigkeit, mit der eine Deletion einer Aminosäure an einer gegebenen Position vorkommt, kann in der Berechnung auch in Faktoren aufgeschlüsselt werden.

**[0027]** Werden die Deletionen bei der Berechnung der Häufigkeit nicht berücksichtigt, kann alternativ erwünscht sein, dass der in die Berechnung eingesetzte Wert für N bei einer gegebenen Aminosäureposition n gleich der Zahl der Varianten vermindert um die Zahl der Varianten ist, bei denen an dieser gegebenen Position ein Freiraum für eine Aminosäure vorkommt. Somit gilt für das erste Beispiel der Aufreihung  $O_{16}^Q/N = 0,333$  (33%) und  $O_{16}^* = 0,667$  (67%), falls der Freiraum für eine Aminosäure als ein Aminosäuretyp definiert ist und falls er es nicht ist gilt  $O_{16}^Q/N = 1,0$ (100%).

**[0028]** Basierend auf dieser Bestimmung für die Häufigkeit des Auftretens von Aminosäuretypen an jeder Position n in der Gesamtzahl der Varianten wird für den Satz von Polypeptid-Antigenen ein "Schwellenwert" für den Einbau eines besonderen Aminosäuretyps an der entsprechenden Position n ermittelt. Es lässt sich eine degenerierte Oligonucleotid-Sequenz erzeugen. Die degenerierte Oligonucleotid-Sequenz ist so entworfen, dass sie an jeder Codonposition die kleinste Zahl von Nucleotidkombinationen aufweist, um für jeden ausgewählten Aminosäuretyp auf Basis des gewählten Schwellenwerts Codons zu erzeugen.

**[0029]** Wenn daher die Population von N Varianten durch die allgemeine Formel

$$A_1 A_2 A_3 A_{n-2} A_{n-1} A_n$$

ausgedrückt wird, in welcher jede Variable  $A_n$  eine an der n-ten Aminosäureposition des Proteins auftretende Aminosäure wiedergibt, lässt sich ein aus der degenerierten Oligonucleotidsequenz gewonnener Satz von Polypeptid-Antigenen durch die allgemeine Formel

$$A'_1 A'_2 A'_3 A'_{n-2} A'_{n-1} A'_n$$

ausdrücken, in welcher jede Variable  $A'_n$  einen Aminosäuretyp darstellt, der von einer möglichen Nucleotid-Kombination an der entsprechenden Codonposition n in der degenerierten Oligonucleotidsequenz codiert wird.

**[0030]** Die eingesetzte Schwellenhäufigkeit zum Auswählen der Aminosäuretypen für den Einbau in den Satz von Polypeptid-Antigenen und folglich das Festlegen der degenerierten Oligonucleotidsequenz kann für jede Aminosäureposition angewendet werden. Beispielsweise kann ein Schwellenwert von 15% über die gesamte Proteinsequenz zur Anwendung kommen. Alternativ kann der Schwellenwert für jede Aminosäureposition n unabhängig festgesetzt werden. Der Schwellenwert kann z.B. an jeder Aminosäureposition n so festgelegt werden, dass die meisten gewöhnlich vorkommenden Typen von Aminosäuren enthalten sind, z.B. solche, die an dieser Position in mindesten 90% der N Varianten vorkommen.

**[0031]** Es kann in manchen Fällen erwünscht sein, ein weiteres Kriterium an die Bestimmung einer degenerierten Oligonucleotidsequenz anzulegen, wobei die Degeneration einer Codonposition so eingeschränkt wird, dass nicht mehr als eine vorgegebene Zahl von Aminosäuretypen in dem Satz der Polypeptid-Antigene an der entsprechenden Position für die Aminosäure auftreten kann. Beispielsweise kann die degenerierte Sequenz an einer vorgegebenen Codonposition n so eingeschränkt werden, dass bei mindestens etwa 11 % der Polypeptide in dem Satz von Polypeptid-Antigenen die ausgewählten Aminosäuren vorkommen. Dies bedeutet, dass von allen möglichen Nucleotidkombinationen für dieses degenerierte Codon nicht mehr als 9 verschiedenen Aminosäuren an dieser Position auftreten. Somit hängt die Häufigkeit, mit der eine besondere Aminosäure

an einer vorgegebenen Position vorkommt von der möglichen Degeneration der entsprechenden Codonposition ab. Vorzugsweise ist die Zahl 11,1 (9 unterschiedliche Aminosäuren), 12,5 (8 verschiedene Aminosäuren), 16,6 (6 verschiedene Aminosäuren), 25 (4 verschiedene Aminosäuren) oder 50 (2 verschiedene Aminosäuren).

**[0032]** Die für die Auswahl der Population von Varianten für eine Analyse der Häufigkeit verwendeten Kriterien können auch mit solchen Faktoren wie dem erwarteten Nutzen des Satzes aus Polypeptid-Antigenen und Faktoren in Bezug auf eine Schutzimpfung oder Tolerierung bestimmt werden. Die Analyse einer variierenden Proteinsequenz kann z.B. auf Subpopulationen einer größeren Population von Varianten des Proteins beschränkt werden, die auf Faktoren wie z.B. epidemiologischen Daten einschließlich dem geographischen Auftreten oder auf bekannten Allelfamilien (wie z.B. Varianten des DQ-HLA-Klasse II-Allels) basieren. Im Falle von Proteinzusammensetzungen von Pathogenen, kann die Population von für die Analyse ausgewählten Varianten auf der Grundlage bekannter Tropismen für einen besonders empfänglichen Wirtsorganismus ausgewählt werden.

**[0033]** Es gibt viele Möglichkeiten, mit welchen sich der Satz von Polypeptid-Antigenen aus der degenerierten Oligonucleotidsequenz erzeugen lässt. Die chemische Synthese eines degenerierten Oligonucleotids lässt sich in einem automatischen DNA-Synthesizer ausführen und die synthetischen Oligonucleotide können dann zur Expression in ein geeignetes Gen ligiert werden. Falls erwünscht kann in die Sequenz ein Startcodon (ATG) eingearbeitet werden. Die degenerierten Oligonucleotidsequenzen können in ein Genkonstrukt eingebaut werden, damit ein im Wesentlichen aus dem Satz von Polypeptid-Antigenen bestehendes Protein exprimiert werden kann. Alternativ kann der Satz von Polypeptid-Antigenen als Teil von Fusionsproteinen exprimiert werden. Die erhaltene Genbibliothek kann durch Manipulation von regulatorischen Transkriptionssequenzen unter eine geeignete Transkriptionskontrolle gebracht werden. Es kann erwünscht sein, Fusionsproteine mit einer Leadersequenz herzustellen, welche den Transport der rekombinanten Proteine auf geeigneten Ausscheidungswegen der Zelle lenkt.

**[0034]** Es sind verschiedene Verfahren zur chemischen Synthese von Polydesoxynucleotiden bekannt, welche wie bei der Peptidsynthese in im Handel erhältlichen DNA-Synthesizern völlig automatisiert worden sind (siehe US-Patent 4,598,049 von Itakura et al., US-Patent 4,458,066 von Caruthers et al. sowie die US-Patente 4,401,796 und 4,373,071 von Itakura).

**[0035]** Der Zweck eines degenerierten Satzes von Oligonucleotiden besteht darin, in einer Mischung alle die Sequenzen zur Verfügung zu stellen, welche den gewünschten Satz von Oligopeptid-Antigenen codieren. Es ist im Allgemeinen unpraktisch, die Oligonucleotide dieser Mischung jeweils nacheinander zu synthetisieren, insbesondere im Falle einer großen Zahl von möglichen Varianten. In diesen Fällen lässt sich die Mischung nach einer Strategie synthetisieren, in welcher eine Mischung von Kopplungseinheiten (Nucleotid-Monomere) an den geeigneten Positionen in der Sequenz zugesetzt werden, so dass das Endnucleotid die Sequenzen umfasst, welche den gewünschten Satz von Oligopeptid-Antigenen codieren. In herkömmlichen Techniken der DNA-Synthese werden vorteilhafterweise an den reaktiven Desoxynucleotiden Schutzgruppen angebracht, so dass nach dem Einbau in ein wachsendes Oligomer ein weiteres Ankoppeln an dieses Oligomer verhindert wird, bis die Schutzgruppe in einem weiteren Schritt wieder entfernt wird. Um eine degenerierte Sequenz zu erhalten, kann während einer Kopplungsrunde mehr als ein Typ von Desoxynucleotid gleichzeitig mit dem wachsenden Oligonucleotid umgesetzt werden, entweder indem die Nucleotide vorher gemischt werden oder indem der Synthesizer so programmiert wird, dass geeignete Volumina von den das Nucleotid enthaltenden Reaktionslösungen angeboten werden. Für jede einer Aminosäureposition mit nur einem Aminosäuretyp in dem eventuellen Satz von Polypeptid-Antigenen entsprechenden Codonposition weist jedes Oligonucleotid des degenerierten Satzes von Oligonucleotiden eine identische Nucleotidsequenz auf. An einer Codonposition, die einer Aminosäureposition entspricht, an welcher im eventuellen Satz mehr als ein Typ von Aminosäure vorkommt, umfasst der degenerierte Satz von Oligonucleotiden Nucleotidsequenzen, welche zu Codons führen, die an dieser Position im Satz solche Typen von Aminosäuren codieren. Wegen anderer Kombinationen, welche die degenerierte Nucleotidsequenz aufweisen kann, weisen in diesen Fällen die erhaltenen Oligonucleotide Codons auf, welche auf andere Typen von Aminosäuren gerichtet sind als die, deren Vorkommen auf der Grundlage der Analyse der Häufigkeit des Auftretens in der Variante beabsichtigt war. Die Synthese von degenerierten Oligonucleotiden ist im Stand der Technik gut bekannt (siehe z.B. Narang, SA (1983) tetrahedron 39:3; Itakura et al. (1981 in Recombinant DNA, Proc.3rd Cleveland Sympos. Macromolecules, Hrg. AG Walton, Amsterdam: Elsevier SS. 273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53: 322; Itakura et al. (1984) Science 198: 1056 und Ike et al. (1983) Nucl. Acid Res. 11:477).

**[0036]** Wie im Stand der Technik bekannt und zur weiteren Veranschaulichung dieser Technik legen Gene, die Proteine codieren, die Aminosäuresequenz über die Reihenfolge der Desoxyribonucleotide in der DaNA

fest, jedoch direkter über die Sequenz der Ribonucleotide in ihren mRNA-Transkripten. Es ist eine wichtige Eigenschaft des genetischen Codes, dass mit Ausnahme von zwei Aminosäuren alle anderen von mehr als einem Nucleotidtriplett (Codon) codiert werden. Der genetische Code (in Form der Desoxyribonucleotide) lässt sich wie folgt darstellen:

TABELLE 1

erste Position (5')	zweite Position				dritte Position (3')
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stp	Stp	A
	Leu	Ser	Stp	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

\*Stp = Stopcodon

**[0037]** Wie oben angemerkt spielt bei einer Strategie zur Synthese von degeneriertem Oligonucleotid während einer vorgegebenen Kopplungsrunde die simultane Reaktion von mehr als einer Art von Desoxynucleotiden eine Rolle. Sollte z.B. entweder ein Histidin (His) oder Threonin (Thr) an einer vorbestimmten Aminosäureposition auftreten, könnte die Synthese des Satzes von Oligonucleotiden wie folgt ausgeführt werden: (unter der Annahme, dass die Synthese von 3' nach 5' fortschreitet) würde das wachsende Oligonucleotid zuerst an ein 5'-geschütztes Thymidin-Desoxynucleotid angekoppelt, die Schutzgruppe entfernt werden und dann simultan mit einer Mischung von 5'-geschütztem Adenin-Desoxynucleotid und einem 5'-geschütztem Cytidin-Desoxynucleotid umgesetzt werden. Nach Entfernung der Schutzgruppen von den erhaltenen Oligonucleotiden wird eine weitere Mischung aus 5'-geschütztem Adenin-Desoxynucleotid und 5'-geschütztem Cytidin-Desoxynucleotid simultan reagieren gelassen. Der erhaltene Satz von Oligonucleotiden weist an dieser Codonposition



entweder ACT (Thr), AAT (Asn), CAT (His) oder CCT (Pro) auf. Wenn mehr als ein Nucleotid eines Codons verändert wird, kann die Verwendung von Nucleotid-Monomeren in der Synthese somit potentiell zu einer Mischung von Codons führen, einschließlich jedoch nicht ausschließlich zu solchen, deren Vorkommen durch die Häufigkeitsanalyse beabsichtigt ist.

**[0038]** Tabelle 1 kann verwendet werden, um degenerierte Nucleotidsequenzen mit möglichen Kombinationen zu berechnen, die Codons für vorgegebene Aminosäuren entsprechen, so dass die Anzahl der Aminosäuretypen an Positionen mit degenerierten Codons, die über der Zahl der mittels Häufigkeitsanalyse ausgewählten liegt, möglichst klein gehalten wird. Zur Verwendung bei der Angabe von degenerierten Oligonucleotidsequenzen sind die folgenden IUPAC-Zeichen und Bedeutungen vorgesehen:

TABELLE 2

Zeichen	Entsprechung
A	A: Adenin
C	C: Cytosin
G	G: Guanin
T	T: Thymin
M	A oder C
R	A oder G
W	A oder T
S	C oder G
Y	C oder T
K	G oder T
V	A oder C oder G; nicht T
H	A oder G oder T; nicht C
D	A oder G oder T; nicht C
B	C oder G oder T; nicht A
N	A oder C oder G oder T oder unbekannt

**[0039]** Um an einer vorgegebenen Aminosäureposition einen Aminosäure-Freiraum (Deletion) zu schaffen, kann während der passenden Runden bei der Ankopplung von Nucleinsäuren (d.h. drei Kopplungsrunden pro Codon) ein Teil der Oligonucleotid-Mischung bei Seite gehalten werden, so dass insgesamt eine bestimmte Codonposition fehlt, und kann dann wieder beim Synthesestart der nächstfolgenden Codonposition der Mischung zurückgesetzt werden.

**[0040]** Nach diesem Verfahren lässt sich die gesamte codierende Sequenz für den Satz von Polypeptid-Antigenen synthetisieren. In einigen Fällen kann es erwünscht sein, mit dieser Methode degenerierte Oligonucleotid-Fragmente zu synthetisieren, die dann an Invariante DNA-Sequenzen ligiert werden, die getrennt synthetisiert wurden, um ein längeres degeneriertes Oligonucleotid herzustellen.

**[0041]** Auch brauchen die Aminosäurepositionen, die in dem hergestellten Satz von Polypeptid-Antigenen mehr als einen Aminosäuretyp aufweisen, in der Polypeptid-Sequenz nicht nebeneinander zu liegen. In einigen Fällen kann es erwünscht sein, eine Anzahl von degenerierten Oligonucleotid-Fragmenten zu synthetisieren, wobei jedes Fragment einem bestimmten Fragment der codierenden Sequenz für den Satz von Polypeptid-Antigenen entspricht. Jedes degenerierte Oligonucleotid-Fragment kann dann enzymatisch an die passenden invarianten DNA-Sequenzen ligiert werden, die Abfolgen von Aminosäuren codieren, von denen an jeder Position im Satz der Polypeptid-Antigene nur ein Aminosäuretyp vorkommt. Somit wird die fertige degenerierte codierende Sequenz durch Fusion von sowohl degenerierten als auch Invarianten Sequenzen gewonnen.

**[0042]** Diese Verfahren sind nützlich, wenn die auf Häufigkeit beruhenden Mutationen in Abschnitten auf dem herzustellenden Polypeptid-Antigen konzentriert sind und es ist wünschenswert, die langen Invarianten Nucleotid-Sequenzen getrennt von der Synthese der degenerierten Nucleotid-Sequenzen zu synthetisieren.

**[0043]** Ferner können die degenerierten Oligonucleotide als degenerierte Fragmente synthetisiert werden und dann aneinander ligiert werden (d.h. es können komplementäre Überhänge geschaffen oder eine Blunt-End-Ligation eingesetzt werden). Gewöhnlich werden überlappende Fragmente als Komplementärstränge synthetisiert und dann werden die verbleibenden einzelsträngigen Abschnitte jedes Stranges hybridisiert und aufgefüllt. In Fällen wo komplementäre Stränge hybridisiert werden müssen, ist es im Allgemeinen wünschenswert, dass die Verbindung in einem Bereich mit geringer Degeneration erfolgt.

**[0044]** Die Nucleotid-Sequenzen, die aus der Synthese einer degenerierten Oligonucleotid-Sequenz stammen und den Satz von Polypeptid-Antigenen codieren können eingesetzt werden, um den Satz von Polypeptid-Antigenen über mikrobielle Prozesse herzustellen. Die Ligation der Sequenzen in ein Genkonstrukt, wie z.B. einen Expressionsvektor, und die Transformation oder Transfektion in entweder eukaryotische Wirte (Hefe, Affen oder Säuger) oder prokaryotische Wirte (Bakterienzellen) stellen bei der Herstellung von anderen gut bekannten Proteinen, wie z.B. Insulin, Interferonen, menschlichen Wachstumshormonen, IL-1, IL-2 und dergl., eingesetzte Standardverfahren dar. Ähnliche Verfahren oder offenkundige Modifikationen derselben können Verwendung finden, um den Satz von Polypeptid-Antigenen mit mikrobiologischen Mitteln oder mittels Gewebekultur-Technologie in Übereinstimmung mit der vorliegenden Erfindung herzustellen.

**[0045]** Wie oben angegeben kann der degenerierte Satz von Oligonucleotiden, der den Satz von Polypeptid-Antigenen in Form einer Bibliothek von Genkonstrukten codiert in einen Vektor ligiert werden, der für eine Expression entweder in prokaryotischen Zellen, oder in eukaryotischen Zellen oder in beiden geeignet ist. Expressionsvehikel zur Herstellung des Satzes von Polypeptid-Antigenen dieser Erfindung umfassen Plasmide oder andere Vektoren. Beispielsweise umfassen geeignete Vektoren für die Expression des degenerierten Satzes von Oligonucleotiden Plasmide der Typen pBR322, pEMBL-Plasmide, pEX-Plasmide, pBTac-Plasmide und pUC-Plasmide für die Expression in prokaryotischen Zellen, wie z.B. *E. coli*.

**[0046]** Es gibt eine Anzahl von Vektoren für die Expression von rekombinanten Proteinen in Hefe. YEP24, YIP5, YEP51, YEP52 und YRP17 sind z.B. Klonierungs- und Expressionsvektoren, die für den Einbau genetischer Konstrukte in *S. cerevisiae* geeignet sind (siehe z.B. Broach et al. (1983) in *Experimental Manipulation of Gene Expression*, Hrg, M. Inouye, Academic Press, S. 83). Diese Vektoren können sich wegen des Vorkommens von pBR322 ori in *E. coli* und in *S. cerevisiae* wegen der Replikations-Determinante des Yeast 2 Micron-Plamids replizieren. Zusätzlich können Marker für eine Arzneimittel-Resistenz, wie z.B. Ampicillin, eingesetzt werden.

**[0047]** Die bevorzugten Expressionsvektoren aus Säugern enthalten sowohl prokaryotische Sequenzen, um die Propagation des Vektors in Bakterien zu erleichtern, als auch eine oder mehrere eukaryotische Transkriptionseinheiten, die in eukaryotischen Zellen exprimiert werden. Die Vektoren pSV2gpt, pSV2neo, PsV2dhfr, pTK2, pRSVneo, pMSG, pSVT7, pko-neo und pHyg sind Beispiele für Expressionsvektoren aus Säugern, die sich für eine Transfektion von eukaryotischen Zellen eignen. Diese Vektoren werden mit Sequenzen aus bakteriellen Plasmiden, wie z.B. pBR322, modifiziert, um die Replikation und die Selektion mittels Arzneimittelresistenz sowohl in prokaryotischen als auch in eukaryotischen Zellen zu erleichtern. Alternativ können Derivate von Viren, wie z.B. das Papilloma-Virus des Rindes (BPV-1) und das Epstein-Barr-Virus (pHEBo und p205) für eine transiente Expression von Proteinen in eukaryotische Zellen eingesetzt werden. Die bei der Präparation der Plasmide und der Transformation von Wirtsorganismen verwendeten verschiedenen Verfahren sind im Stand der Technik gut bekannt. Zu anderen bekannten sowohl prokaryotischen als auch eukaryotischen Expressionssystemen sowie zu allgemeinen rekombinanten Verfahren siehe *Molecular Cloning*, 2. Ausgabe, herausgegeben von Sambrook, Fritsch und Maniatis (Cold Spring Harbor Laboratory Press: 1989).

**[0048]** Zur Expression der Bibliothek von Genkonstrukten des degenerierten Satzes von Oligonucleotiden kann es erwünscht sein, regulatorische Transkriptions- und Translationselemente sowie andere nicht codierende Sequenzen in das Expressionskonstrukt einzubauen. Es können z.B. regulatorische Elemente mit konstitutiven und induzierbaren Promotoren und Enhancern eingebaut werden.

**[0049]** In einigen Fällen ist es notwendig, ein Start-Codon (ATG) an die degenerierte Oligonucleotidsequenz anzuhängen. Im Stand der Technik ist bekannt, dass Methionin an der N-terminalen Position unter Einsatz des Enzyms Methioninaminopeptidase (MAP) enzymatisch abgespalten werden kann. MAP ist aus *E. coli* (Ben Bassat et al. (1987) *J. Bacteriol.* 169:151-757) und *Salmonella typhimurium* kloniert worden und seine in vitro-Aktivität an rekombinanten Proteinen nachgewiesen worden (Miller et al. (1987) *PNAS* 84:2718-1722). Daher lässt sich die Entfernung eines N-terminalen Methionins, falls erwünscht, entweder in vivo durch Expression des Satzes von Polypeptid-Antigenen in einem MAP produzierenden Wirt (z.B. *E. coli* oder *S. cerevisiae*) oder in vitro durch Einsatz von gereinigtem MPA (z.B. Verfahren von Miller et al.) erzielen.

**[0050]** Alternativ können die die Polypeptid-Antigene codierenden Sequenzen als ein Teil eines Fusionsgens eingebaut werden, das ein endogenes Protein für die Expression durch den Mikroorganismus enthält. Beispielsweise kann das VP6-Capsid-Protein des Rotavirus entweder in der monomeren Form oder in Form eines Virusteilchens als immunologisches Trägerprotein für den Satz von Polypeptid-Antigenen eingesetzt werden. Der Satz von degenerierten Oligonucleotid-Sequenzen kann in das Konstrukt eines Fusionsgens eingebaut werden, das codierende Sequenzen für ein spätes Strukturprotein eines Vaccinia-Virus enthält, um einen Satz

von rekombinaten Viren herzustellen, welche Fusionsproteine mit dem Satz von Polypeptid-Antigenen als Teil des Virions exprimieren. Es ist mit Hilfe von V3-Loop/Hepatitis B-Oberflächenantigen-Fusionsproteinen gezeigt worden, dass rekombinante Hepatitis B-Virionen für diese Rolle ebenso verwendet werden können. Auf ähnliche Weise lassen sich chimäre Konstrukte herstellen, welche die Fusionsproteine codieren, die den Satz von Polypeptidantigenen sowie das Capsid-Protein des Poliovirus enthalten, um die Immunogenität des Satzes von Polypeptid-Antigenen zu steigern. Der Einsatz eines solchen Expressionssystems für die Fusionsproteine zur Herstellung eines Satzes von Polypeptid-Antigenen weist den Vorteil auf, dass oft eine B-Zell-Proliferation als Antwort auf das Immunogen hervorgerufen werden kann (siehe z.B. EP-Veröffentlichungsnummer 0259149 sowie Evans et al. (1989) *Nature* 339:385; Huang et al. (1988) *J. Virolog.* 62:3855 und Schlienger et al. (1992) *J. Virolog.* 66:2). Es kann das Multiple-Antigen-Peptid (MAP)-System für auf Peptiden basierende Vakzine eingesetzt werden, in welchem der Satz von Polypeptid-Antigenen direkt aus der organisch-chemischen Synthese der Peptide auf einem oligomeren verzweigenden Lysinkern erhalten wird (siehe z.B. Posnett et al. (1988) *JBC* 263:1719 und Nardelli et al. (1992) *J. Immunol.* 148:914). Es können auch fremde Antigen-determinanten von Bakterienzellen exprimiert und präsentiert werden.

**[0051]** Techniken zur Gewinnung von Fusionsgenen sind gut bekannt. Das Zusammenfügen von zahlreichen DNA-Fragmenten, die unterschiedliche Polypeptid-Sequenzen codieren, wird mit konventionellen Techniken durchgeführt, indem für die Ligation glatte oder versetzt angeordnete Enden, für passende Enden ein Verdau mit Restriktionsenzymen, ein zweckmäßiges Auffüllen der kohäsiven Enden, zur Vermeidung von unerwünschten Verknüpfungen eine Behandlung mit alkalischer Phosphatase sowie eine enzymatische Ligation zum Einsatz kommen. Alternativ lassen sich die Fusionsgene mit konventionellen Techniken einschließlich einem automatischen DNA-Synthesizer synthetisieren.

**[0052]** Eine alternative Lösungsmöglichkeit zur Gewinnung des Satzes von Polypeptid-Antigenen besteht darin, die Synthese der Peptide direkt auszuführen. An jeder Codonposition  $n$  im degenerierten Oligonucleotid kann jede mögliche Kombination von Nucleotiden festgelegt und die entsprechende Aminosäure zum Einbau an der entsprechenden Aminosäureposition des Satzes von Polypeptid-Antigenen angegeben werden. Somit lässt sich eine Synthese einer degenerierten Polypeptidsequenz ausrichten, in welcher Sequenz an den Aminosäurepositionen eine Divergenz auftritt, an welchen in der entsprechenden Codonposition des degenerierten Oligonucleotids mehr als eine Aminosäure codiert wird. Die organisch-chemische Synthese von Polypeptiden ist gut bekannt und kann mit Verfahren wie der Festphasen-Peptidsynthese unter Einsatz von automatischen Protein-Synthesizern durchgeführt werden.

**[0053]** Die Synthese von Polypeptiden erfolgt im Allgemeinen über die Kondensation der Carboxygruppe einer Aminosäure und der Aminogruppe einer anderen Aminosäure, um eine Peptidbindung zu bilden. Eine Sequenz kann aufgebaut werden, indem in schrittweiser Verlängerung die Kondensation von einzelnen Aminosäureresten auf eine Weise erfolgt, die der Synthese von Oligonucleotiden analog ist. Bei solchen Kondensationen können die Amino- und Carboxygruppen, die nicht an der Reaktion teilnehmen sollen, mit Hilfe von Schutzgruppen blockiert werden, die leicht eingeführt werden können, stabil gegenüber den Kondensationsreaktionen sind und selektiv aus dem fertigen Peptid entfernt werden können. Somit umfasst das gesamte Verfahren im Allgemeinen einen Schutz, einer Aktivierung, eine Kopplung sowie die Entfernung der Schutzgruppen Falls in einem Peptid Aminosäuren mit Seitenketten vorkommen, die während der Kondensation reagieren können, können die Seitenketten ebenfalls reversibel geschützt werden und die Schutzgruppen im abschließenden Schritt der Synthese entfernt werden.

**[0054]** Bei einer erfolgreichen Synthese von großen Polypeptiden mittels einer linearen Strategie müssen bei jedem chemischen Schritt nahezu quantitative Ausbeuten erzielt werden. Bei vielen automatischen Syntheseverfahren für Peptide wird der Vorteil ausgenutzt, die wachsende Polypeptidkette an einen unlöslichen Träger aus Polymerharz anzuheften, so dass das Polypeptid nach jedem Reaktionsschritt von Nebenprodukten und überschüssigen Reaktanten frei gewaschen werden kann (siehe z.B. Merrifield (1963) *J.A.C.S.* 85:2149; Chang et al. (1978) *Int. J. Peptide Protein Res.* 11: 246; Barany und Merrifield *Ther Peptides* Bd.2 ©1979 NY: Academic Press, SS. 1–284; Tam, J.P. (1988) *PNAS* 85: 5409 und Tam et al. US-Patent 4,507,230). Eine erste Aminosäure wird z.B. über eine abspaltbare Bindung an ihrer Carbocylgruppe an ein Harz angeheftet, die Blockierung an der Aminogruppe entfernt und mit einer zweiten aktivierten Aminosäure mit einer geschützten  $\alpha$ -Aminogruppe verbunden. Von dem erhaltenen geschützten Dipeptid wird die Schutzgruppe entfernt, um ein freies Aminoende zu erhalten und das nun ungeschützte Dipeptid mit einer dritten N-geschützten Aminosäure verbunden. Nach vielen Wiederholungen dieser Schritte wird das fertige Polypeptid von dem Träger aus Harz abgespalten und die Schutzgruppen auf geeignete Weise entfernt.

**[0055]** Zur Herstellung des Satzes von Polypeptid-Antigenen kann mehr als ein N-geschützter Aminosäuretyp

gleichzeitig in jeder Kopplungsrunde mit der wachsenden Polypeptidkette umgesetzt werden, um an jeder Aminosäureposition die gewünschte degenerierte Aminosäuresequenz zu erhalten. In einer Ausführungsform umfasst der Satz von Polypeptiden nur solche Aminosäuren, die an jeder Position  $n$  in der Population von Varianten oberhalb der vorausbestimmten Schwellenhäufigkeit vorkommen. Alternativ kann man zuerst das degenerierte Oligonucleotid festlegen, die von der Kombination von Codonen codierten Aminosäuren ermitteln und alle diese Aminosäuren bei der chemischen Synthese einsetzen. Beispielsweise lässt sich ein degeneriertes Codon an der Codonposition  $n$  mit der Sequenz MMT, das dann folglich entweder Thr (ACT), Asn (AAT), His (CAT) oder Pro (CCT) codiert, im Zuge der Peptidsynthese herstellen, indem alle vier N-geschützten Aminosäuretypen gleichzeitig mit dem freien Aminoende des wachsenden, an das Harz gebundenen Peptids reagieren gelassen werden. Somit werden vier Peptid-Subpopulationen erzeugt, wobei jede Subpopulation durch den Aminosäuretyp definiert wird, der an der Aminosäureposition  $n$ , die der Codonposition  $n$  entspricht, vorkommt.

**[0056]** Weil die dem an das Harz gebundenen Polypeptid zugesetzte Aminosäure geschützt ist, wird das Wachstum der Peptidkette nach der Zugabe der geschützten Aminosäure bis zu dem nachfolgenden Schritt der Entfernung der Schutzgruppe beendet. Der Fachmann wird erkennen, dass es wegen der potentiellen Unterschiede in der Reaktivität der zahlreichen Aminosäureanaloge erwünscht sein kann, nicht äquimolare Verhältnisse der Aminosäuretypen einzusetzen, wenn gleichzeitig mehr als ein Aminosäuretyp reagieren gelassen wird, um bei den Subpopulationen äquimolare Verhältnisse zu erzielen. Alternativ kann es erwünscht sein, das an das Harz gebundene Polypeptid in Aliquots aufzuteilen, von denen jedes mit einem bestimmten Aminosäuretyp reagieren gelassen wird, wobei die Polypeptidprodukte dann vor der nächsten Kopplungsreaktion wieder vereinigt werden. Diese Technik kann angewandt werden, um in einer Subpopulation eine Aminosäurelücke zu erzeugen, einfach indem während einer Kopplungsrunde ein passendes Aliquot bei Seite gehalten wird und dann vor der nächsten Kopplungsrunde alle an das Harz gebundenen Polypeptide wieder vereinigt werden. Ferner ist leicht einzusehen, dass bei den vielen zur Verfügung stehenden unterschiedlichen Schutz- und Aktivierungsgruppen die chemische Synthese des Polypeptids entweder in Richtung N-Terminus zu C-Terminus oder in Richtung C-Terminus zu N-Terminus erfolgen kann.

**[0057]** Der erzeugte Satz von Polypeptid-Antigenen kann mit proteinfreien Stoffen wie z.B. Lipiden oder Kohlenhydraten kovalent oder nicht kovalent modifiziert werden, um die Immunogenität oder Löslichkeit zu steigern. Die vorliegende Erfindung soll alle derartigen chemischen Modifikationen des Satzes von Polypeptid-Antigenen mit umfassen, solange die modifizierten Peptid-Antigene im Wesentlichen über alle antigenen/immunogenen Eigenschaften der Ausgangsmischung verfügen.

**[0058]** Der erzeugte Satz von Polypeptid-Antigenen kann auch an ein Virusteilchen, ein replizierendes Virus oder an andere Mikroorganismen angekoppelt oder in diese eingebaut werden, um die Immunogenität zu steigern. Der Satz von Polypeptid-Antigenen kann auf chemischem Wege an das Virusteilchen oder den Mikroorganismus oder einen immunogenen Teil davon angeheftet werden.

**[0059]** Es gibt eine große Zahl von chemischen Vernetzungsmitteln, die dem Fachmann bekannt sind. Für die vorliegende Erfindung sind die bevorzugten Vernetzungsmittel heterobifunktionelle Vernetzungsmittel, die dafür verwendet werden können, Proteine schrittweise aneinander zu binden. Heterobifunktionelle Vernetzungsmittel sorgen für die Möglichkeit, spezifischere Kopplungsverfahren für die Konjugation von Proteinen zu entwerfen, wobei das Auftreten unerwünschter Nebenreaktionen, wie z.B. von Polymeren, aus Homoproteinen verringert wird. Im Stand der Technik ist eine große Vielfalt von heterobifunktionellen Vernetzungsmitteln bekannt. Diese umfassen Succinimidyl-4(N-maleimidomethyl)cyclohexan-1-carboxylat (SMCC), m-Maleimido-benzoyl-N-hydroxysuccinimidester (MBS); N-Succinimidyl (4-iodoacetyl)-aminobenzoat (SIAB), Succinimidyl-4-(p-maleimidophenyl)butyrat (SMPB), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimidhydrochlorid (EDC); 4-Succinimidylloxycarbonyl- $\alpha$ -methyl- $\alpha$ -(2-pyridyldithio)toluen (SMPT), N-Succinimidyl-3-(2-pyridyldithio)propionat (SPDP), Succinimidyl 6-[3-(2-pyridyldithio)propionate]hexanoat (LC-SPDP). Diese Quervernetzungsmittel mit N-hydroxysuccinimid-Resten können als N-Hydroxysulfosuccinimid-Analoga erhalten werden, die im Allgemeinen eine größere Löslichkeit in Wasser aufweisen. Zusätzlich lassen sich solche Quervernetzungsmittel mit Disulfidbrücken innerhalb der Verbindungskette an Stelle der Alkylderivate synthetisieren, so dass in vivo die Menge der Linker-Abspaltungen verringert wird.

**[0060]** Das Einführen eines Antigens in ein Tier löst eine Reihe von Ereignissen aus, die in sowohl zellulärer als auch humoraler Immunität kulminieren. Nach Konvention wird die Eigenschaft eines Moleküls, eine Immunantwort induzieren zu können als Immunogenität bezeichnet. Die Eigenschaft, mit einem Antikörper, der eingeführt worden war reagieren zu können, wird als Antigenität bezeichnet. Antikörper, die befähigt sind, mit zwei oder mehr unterschiedlichen Antigenen eine Quervernetzung einzugehen, können dies auf Grund eines ge-

wissen Grades von struktureller und chemischer Ähnlichkeit zwischen den Antigen-determinanten (oder "Epi-topen") der Antigene. Ein Immunogen-Protein setzt sich gewöhnlich aus einer Anzahl von Antigen-determinanten zusammen. Daher führt die Immunisierung mit einem Protein zu der Bildung von Antikörpermolekülen mit unterschiedlichen Spezifitäten, wobei die Anzahl der unterschiedlichen Antikörper von der Anzahl der Antigen-determinanten und deren inhärenter Immunogenität abhängt.

**[0061]** Proteine sind hoch immunogen, wenn sie in ein Tier injiziert werden, für welches sie keine normalen ("Selbst-") Bestandteile darstellen. Umgekehrt lösen Peptide und andere Verbindungen mit Molekulargewichten unter ca. 5000 Dalton ("Haptene" genannt) von sich aus im Allgemeinen keine Bildung von Antikörpern aus. Werden diese kleinen Molekülantigene jedoch zuerst an ein langes immunogenes Antigen wie z.B. ein Protein gekoppelt, können Antikörper gebildet werden, die sich spezifisch an die Epitope auf den kleinen Molekülen binden. Eine Konjugation von Haptenen an Trägerproteine kann wie oben beschrieben erfolgen.

**[0062]** Wenn nötig sollte bei der Modifizierung eines solchen Liganden zur Herstellung eines Immunogens die Wirkung auf die strukturelle Spezifität des Antikörpers in Betracht gezogen werden. Das heißt, indem eine Stelle auf dem Liganden zur Konjugation an einen Träger wie z.B. ein Protein ausgesucht wird und die ausgesuchte Stelle so gewählt wird, dass die Verabreichung des erhaltenen Immunogens zu Antikörpern führt, die den ursprünglichen Liganden erkennen. Ferner muss der Antikörper nicht nur den ursprünglichen Liganden erkennen, sondern es müssen die signifikanten Eigenschaften des Ligandenteils des Immunogens beibehalten werden, so dass der nach Verabreichung des Immunogens gebildete Antikörper mit größerer Wahrscheinlichkeit zwischen Verbindungen unterscheidet, die mit dem Liganden nahe verwandt sind, die ebenfalls in der Probe des Patienten vorkommen können. Darüber hinaus sollten die Antikörper über hohe Bindungskonstanten verfügen.

**[0063]** Vakzine mit dem hergestellten Satz von Polypeptid-Antigenen sowie Varianten derselben mit antigenen Eigenschaften können mit im Stand der Technik bekannten Verfahren hergestellt werden. Solche Vakzine lassen sich z.B. als Injektionsmittel herstellen, z.B. als flüssige Lösungen oder Suspensionen. Es können auch feste Formen zur Auflösung oder Suspension in einer Flüssigkeit vor der Injektion hergestellt werden. Wahlweise kann das Präparat auch emulgiert werden. Der (die) aktive(n) Inhaltsstoff(e) kann (können) mit Trägersubstanzen vermischt werden, die pharmazeutisch verträglich und mit dem aktiven Inhaltsstoff kompatibel sind. Beispiele für geeignete Trägersubstanzen sind Wasser, Saline, Dextrose, Glycerin, Ethanol oder dergl. sowie Kombinationen derselben. Falls erwünscht kann das Vakzin darüber hinaus noch geringere Mengen an Hilfs-substanzen wie z.B. Benetzungsmittel oder Emulgatoren, pH-Puffer oder Adjuvantien wie z.B. Aluminiumhydroxid oder Muramyldipeptid oder Variationen derselben enthalten. Im Falle von Peptiden steigert das Anknüpfen an größere Moleküle wie z.B. an das Hämocyanin der Napfschnecke (KLH) manchmal die Immunogenität. Herkömmlicherweise werden die Vakzine parenteral oder z.B. mittels Injektion, entweder subkutan oder intramuskulär, verabreicht. Zusätzliche Formulierungen, die für andere Verabreichungsarten geeignet sind, umfassen Suppositorien und in einigen Fällen orale Formulierungen. Für Suppositorien umfassen die traditionellen Bindemittel und Trägersubstanzen z.B. Polyalkylene, Glykole oder Triglyceride. Suppositorien lassen sich aus Mischungen bilden, die den aktiven Inhaltsstoff im Bereich von etwa 0,5% bis etwa 10%, vorzugsweise etwa 1% bis etwa 2% enthalten. Orale Formulierungen können solche normalerweise verwendeten Trägersubstanzen wie z.B. Mannit, Lactose, Stärke, Magnesiumstearat, Natrium-Saccharin, Cellulose, Magnesiumcarbonat und dergl. für pharmazeutische Zwecke enthalten. Die Zusammensetzungen können die Form von Lösungen, Suspensionen, Tabletten, Pillen, Kapseln, Formulierungen für eine Dauerausschüttung oder Pulvern einnehmen und enthalten ca. 10% bis ca. 95%, vorzugsweise ca. 25% bis ca. 70%, des aktiven Inhaltsstoffs.

**[0064]** Die aktiven Verbindungen können in das Vakzin in neutraler Form oder als Salz formuliert werden. Pharmazeutisch verträgliche Salze sind (mit der freien Aminogruppe der Polypeptide gebildete) saure Salze, die sich mit anorganischen Säuren wie z.B. Salzsäure oder Phosphorsäuren oder solchen organischen Säuren wie Essigsäure, Oxalsäure, Traubensäure, Mandelsäure und dergl. bilden. Mit den freien Carboxylsäuren gebildete Salze können auch von anorganischen Basen, wie z.B. Natrium-, Kalium-, Ammonium-, Calcium- oder Eisenhydroxiden und solchen organischen Basen wie z.B. Isopropylamin, Trimethylamin, 2-Ethylaminoethanol, Histidin, Procain und dergl. abstammen. Eine Vakzinzusammensetzung kann Peptide umfassen, die Epitope für T-Helferzellen zusammen mit Proteinfragmenten mit der hauptsächlich neutralisierenden Domäne enthalten. Einige dieser Epitope sind z.B. in der Hülle von HIV kartographisch erfasst worden und es ist gezeigt worden, dass diese Regionen die Proliferation und die Freisetzung von Lymphokinen aus Lymphozyten stimulieren. Wenn ein Vakzin mit einem erzeugten Satz von Polypeptid-Antigenen, die aus der Analyse von HIV-1-Isolaten stammen, mit beiden dieser Epitope versehen wird, kann dies zu einer Stimulation von sowohl humoralen als auch zellulären Immunantworten führen. Zusätzlich stehen im Handel erhältliche Träger und Adjuvantien zur Verfügung, um die Immunmodulation von sowohl B-Zell- als auch T-Zell-Populationen für ein Im-

munogen zu erhöhen (z.B. das Imject Supercarrier™-System, Pierce Chemical, Katalog-Nr. 77151G).

**[0065]** Alternativ kann eine Vakzin-Zusammensetzung eine Verbindung enthalten, welche die allgemeine Immunantwort verstärkt. Eine solche Verbindung ist Interleukin-2 (IL-2), von dem berichtet worden ist, dass es über eine allgemeine Immunstimulation die Immunogenität erhöht (Nunberg et al. (1988) in *New Chemical and Genetic Approaches to Vaccination*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). IL-2 kann an die Polypeptide des hergestellten Satzes von Polypeptid-Antigenen gekoppelt werden, um die Wirksamkeit der Impfung zu erhöhen.

**[0066]** Die Impfstoffe werden auf kompatible Weise mit der Formulierungsdosis und in einer therapeutisch wirksamen und immunogenen Menge verabreicht. Die zu verabreichende Menge hängt von dem zu behandelnden Subjekt, der Fähigkeit des Immunsystems des Subjekts zur Synthese von Antikörpern sowie vom Ausmaß des erwünschten Schutzes ab. Die genauen Mengen des für eine Verabreichung erforderlichen Inhaltsstoffs hängen von der Einschätzung des Arztes ab und sind für jedes Individuum eigentümlich. Geeignete Abläufe für die Verabreichung am Anfang und für Boosterdosen sind ebenfalls variabel, typisch ist jedoch eine anfängliche Verabreichung, der eine nachfolgende Injektion oder eine andere Verabreichung nach einem Zeitraum von einer oder zwei Wochen folgt.

**[0067]** Antigene, die eine Toleranz induzieren werden Toleragene genannt, um sie von den Immunogenen, welche eine Immunität hervorrufen, zu unterscheiden. Wird ein Individuum immunogenen Antigenen ausgesetzt, stimuliert dies eine spezifische Immunität und die meisten immunogenen Proteine erzeugen nach einer Exposition verstärkte sekundäre Reaktionen. Im Gegensatz dazu wird bei einer Exposition mit einem toleragenen Antigen nicht nur keine spezifische Immunität induziert, sondern es wird durch eine nachfolgende Verabreichung der immunogenen Formen des gleichen Antigens auch die Aktivierung der Lymphozyten gehemmt. Je nach der physikalisch-chemischen Form, der Dosis und dem Weg der Verabreichung können viele fremde Antigene Immunogene oder Toleragene sein. Diese Fähigkeit, die Reaktionen auf Antigene zu manipulieren kann klinisch ausgenutzt werden, um eine spezifische Immunität zu verstärken oder zu unterdrücken. Es kann z.B. im Zusammenhang mit der Organtransplantationstechnologie erwünscht sein, den Empfänger eines Transplantats mit einem Satz von Polypeptid-Antigenen zu tolerisieren, der aus der Häufigkeitsanalyse eines bekannten Haplotyps eines Peptids der Klasse II (wie z.B. das DQ- oder DR-Allelprodukt), der auf dem transplantierten Gewebe vorkommt, abgeleitet ist, um eine Abstoßung möglichst klein zu halten. Es liegt auch im Äquivalenzbereich dieser Erfindung, dass der Satz von Polypeptid-Antigenen chemisch an ein Apoptosemittel gekoppelt oder als Teil eines Fusionsproteins mit dem Apoptosemittel eingebaut wird, beispielsweise einem Mittel, das die Deregulation der C-myc-Expression oder eines Zelltoxins wie z.B. des Diphtherie-Toxoids verursacht, so dass in antigenspezifischer Weise der programmierte Zelltod ausgelöst wird.

**[0068]** Es ist somit für einen Fachmann reine Routine, die geeignete Verabreichungsweise zu bestimmen, die notwendig ist, um an dem Satz von Polypeptid-Antigenen der vorliegenden Erfindung eine Toleranz zu induzieren.

**[0069]** Das folgende Beispiel dient der weiteren Veranschaulichung der vorliegenden Erfindung.

**[0070]** Das humane Immunodeficiency-Syndrom-Virus vom Typ I (HIV-1), der Verursacher des Acquired-Immunodeficiency-Syndroms (AIDS), zeigt bei unterschiedlichen Isolaten eine sehr ausgeprägte Diversität in der Sequenz. Es ist gezeigt worden, dass das Glykoprotein gp120 der Außenhülle von HIV-1, welches das Anbinden des Virions an CD4 erleichtert, das Hauptziel der neutralisierenden Antikörper ist. Untersuchungen beim Menschen und der Maus haben einen kleinen Abschnitt dieses Proteins mit dem Namen V3-Loop zwischen den Cysteinresten 303 und 338 enthüllt, der die neutralisierenden Hauptantikörper zu dem Virus lockt. Es ist gezeigt worden, dass rekombinante oder synthetische Polypeptide mit V3-Loop-Epitopen eines Isolats von HIV-1-Viren relativ hohe Titer von stammspezifischen neutralisierenden Antikörpern induziert. In einigen Fällen genügen jedoch schon einzelne Substitutionen von Aminosäuren innerhalb des V3-Loops, um eine Bindung der Antikörper stark zu reduzieren, was mit der strengen Spezifität von neutralisierenden Antikörpern gegenüber dem V3-Loop in Übereinstimmung steht.

**[0071]** In Tabelle 3 werden die Ergebnisse gezeigt, die aus einer Häufigkeitsanalyse von V3-Loop-Sequenzen aus einer Population von HN-1-Isolaten erhalten wurden, welche größtenteils von AIDS-Patienten aus Nordamerika stammen (Wolinsky et al. (1992) *Science* 225: 1134; LaRosa et al. (1990) *Science* 249:931 und Holley et al. (1991) *PNAS* 88:6800). Die Aufreihung für jede Sequenzvariante steht in Beziehung zur Referenzsequenz:

1            10            20            30            40  
 |            |            |            |            |  
 CTRPNNN\*\*TRKSIHI\*\*GPGRIFY\*TTGEIIGDIRQAHC            (Seq ID No. 5)

worin Cys-1 der Referenzsequenz dem Cys-303 von gp120 nach der hier verwendeten Nomenklatur entspricht.

TABELLE 3

## Häufigkeitsanalyse von nordamerikanischen HIV-1-Isolaten

Position von V3	Häufigkeit des Auftretens (ausgedrückt in Prozent)
1	C = 100
2	T = 86,3; I = 11,3; A = < 1; L = < 1; M = < 1; P = < 1; S = < 1
3	R = 99,4; I = < 1; K = < 1
4	P = 97,3; H = 1,5; L = < 1; S = < 1
5	N = 86,3; S = 7,0; Y = 3,0; G = 1,8; D = < 1; H = < 1
6	N = 90,9; S = 3,4; D = 1,5; Y = 1,2; G = < 1; I = < 1; K = < 1; T = < 1
7	N = 92,4; T = 3,0; Y = 12; D = < 1; H = < 1; I = < 1; K = < 1; R = < 1
8	* = 99,8; I = < 2; K = < 1
9	* = 99,7; K = < 1
10	T = 92,0; I = 1,8; V = 1,8; A = 1,5; K = < 1; R = < 1
11	R = 88,9; K = 6,3; I = 1,8; E = < 1; G = < 1; M = < 1; P = < 1; Q = < 1; T = < 1
12	K = 76,5; R = 17,4; Q = 2,8; N = 2,1; H = < 2; E = < 1; G = < 1
13	S = 64,8; G = 21,7; R = 10,6; * = < 1; A = < 1; H = < 1; K = < 1
14	I = 96,1; L = 1,2; E = < 1; F = < 1; M = < 1; T = < 1; V = < 1
15	H = 49,2; N = 9,6; P = 9,0; R = 9,0; T = 8,4; Y = 6,2; S = 5,6; F = 1,5; A = < 1; G = < 1; K = < 1; V = < 1
16	I = 72,4; M = 18,7; L = 2,0; T = 1,7; F = 1,2; V = 1,2; K = < 1; S = < 1; R = < 1; Y = < 1
17	* = 97,3; Q = 2,5; R = < 1
18	* = 97,3; R = 2,5; G = < 1
19	G = 90,6; M = 7,4; A = < 1; E = < 1; I = < 1; R = < 1; T = < 1
20	P = 88,9; G = 7,9; L = 1,2; A = < 1; Q = < 1; S = < 1
21	G = 99,0; E = < 1; R = < 1
22	R = 80,1; K = 11,9; S = 3,7; Q = 2,5; * = < 1; G = < 1; M = < 1
23	A = 86,0; V = 4,4; T = 4,2; K = 1,5; R = 1,5; N = 1,2; P = < 1; S = < 1; W = < 1
24	F = 75,9; W = 8,1; I = 7,1; V = 3,4; L = 2,5; Y = 1,7; S = < 1; T = < 1

25	Y = 84,0; H = 6,4; V = 4,2; L = 2,0; F = 1,2; I = < 1; M = < 1; N = < 1; R = < 1
26	* = 99,4; H = < 1; T = < 1
27	A = 52,4; T = 43,9; V = 1,8; * = < 1; Q = < 1; S = < 1; Y = < 1
28	T = 88,4; I = 3,7; R = 2,7; A = 2,4; Q = 1,2; K = < 1; = < 1; P = < 1; Y = < 1
29	G = 76,8; N = 4,9; T = 2,7; H = 2,1; K = 1,8; R = 1,8; D = 1,2; A = < 1; I = < 1; P = < 1; Q = < 1; * = < 1
30	E = 30,8; D = 23,8; R = 9,82; K = 8,2; * = 7,6; Q = 7,9; N = 3,4; G = 2,7; A = < 1; I = < 1; S = < 1; T = < 1
31	I = 87,5; F = 7,6; V = 2,7; L = < 1; K = < 1; M = < 1; R = < 1
32	I = 77,7; V = 8,8; T = 6,1; * = 87,5; Q = 1,2; R = < 2; A = < 1; E = < 1; G = < 1; K = < 1; L = < 1; M = < 1
33	G = 96,6; E = 1,2; A = < 1; K = < 1; N = < 1; R = < 1; s = < 1
34	D = 84,2; N = 16,4; G = < 1; I = < 1; M = < 1; R = < 1; T = < 1
35	I = 92,1; M = 4,6; F = 2,1; E = < 1; L = < 1; T = < 1
36	R = 96,9; * = < 1; E = < 1; G = < 1; K = < 1; S = < 1; C = < 1
37	Q = 82,6; K = 11,9; R = 3,5
38	A = 100
39	H = 90,6; Y = 4,5; R = 3,3; Q = 1,6
40	C = 99,2; Y = < 1

\* (stellt eine Aminosäure-Lücke dar)

**[0072]** Aus den in Tabelle 3 berechneten Daten für die Häufigkeit des Auftretens haben wir die am häufigsten auftretenden Aminosäuretypen an jeder Position ermittelt, die zusammen in mindestens 90% der analysierten Varianten repräsentiert werden. Für jede Aminosäureposition wurde das entsprechende degenerierte Codon ausgewählt und zur Aufstellung einer degenerierten Oligonucleotidsequenz herangezogen, welche die Codons für die 10 Aminosäurereste enthält, die die Sequenzen für den V3-Loop auf jeder Seite flankieren, und welche durch die allgemeine Sequenz

5'-GTTCAGCTGAACGAATCTGTTGACATCAACTGCAYCCGT  
 CCGARCAACAACACCARAARARGMATCMVCATSGGCCCG  
 GGCARAGYTWTCYACRCTAYCSRGRMWTCRYTGGTRAC  
 ATCCGTMAGGCTCACTGCAACATCTCTCGTGCTAAATGG  
 AACAACT-3' (Seq ID No. 6)

wiedergegeben wird. Dieses degenerierte Nucleotid codiert Polypeptid-Antigene, welche durch die allgemeine Sequenz

Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Xaa<sub>1</sub> Arg Pro Xaa<sub>2</sub> Asn  
 Asn Thr Xaa<sub>3</sub> Xaa<sub>4</sub> Xaa<sub>5</sub> Ile Xaa<sub>6</sub> Xaa<sub>7</sub> Gly Pro Gly Xaa<sub>8</sub> Xaa<sub>9</sub> Xaa<sub>10</sub>  
 Xaa<sub>11</sub> Xaa<sub>12</sub> Xaa<sub>13</sub> Xaa<sub>14</sub> Xaa<sub>15</sub> Xaa<sub>16</sub> Xaa<sub>17</sub> Gly Xaa<sub>18</sub> Ile Arg  
 Xaa<sub>19</sub> Ala His Cys Asn Ile Ser Arg Ala Lys Trp Asn Asn Thr  
 (Seq ID No. 7)

wiedergegeben werden, in welcher

Xaa<sub>1</sub> Thr oder Ile ist  
 Xaa<sub>2</sub> Asn oder Ser ist  
 Xaa<sub>3</sub> Arg oder Lys ist  
 Xaa<sub>4</sub> Arg oder Lys ist



Xaa<sub>5</sub> Arg, Gly oder Ser ist  
 Xaa<sub>6</sub> Asn, Pro, Ser, Arg, Thr oder His ist  
 Xaa<sub>7</sub> Met oder Ile ist  
 Xaa<sub>8</sub> Lys oder Arg ist  
 Xaa<sub>9</sub> Val oder Ala ist  
 Xaa<sub>10</sub> Ile oder Phe ist  
 Xaa<sub>11</sub> His oder Tyr ist,  
 Xaa<sub>12</sub> Ala oder Thr ist,  
 Xaa<sub>13</sub> Ile oder Thr ist,  
 Xaa<sub>14</sub> Arg, Asn, Glu oder Gly ist,  
 Xaa<sub>15</sub> Gly, Arg, His, Gln, Asp oder Glu ist,  
 Xaa<sub>16</sub> Phe oder Ile ist,  
 Xaa<sub>17</sub> Ala, Thr, Val oder Ile ist,  
 Xaa<sub>18</sub> Asn oder Asp ist und  
 Xaa<sub>19</sub> Lys oder Gln ist.

**[0073]** Wie oben beschrieben kann das degenerierte Oligonucleotid mit den geeigneten DNA-Sequenzen enzymatisch ligiert werden, um eine Gen-Bibliothek zu erzeugen, welche die Proteine mit dem Satz von Polypeptid-Antigenen codiert.

**[0074]** Auch die an jeder Position am häufigsten auftretenden Aminosäuretypen, die zusammen in mindestens 80% der Varianten repräsentiert werden, wurden analysiert. In diesem Falle wird die ermittelte degenerierte Oligonucleotidsequenz durch die allgemeine Sequenz

```
5' -GTT CAG CTG AAC GAA TCT GTT GAC ATC AAC TGC ACC CGT
    CCG AAC AAC AAC ACC CGT ARA RGC ATC MVC ATS GGC CCG
    GCC CGT GCT WTC TAC RCT ACC GRG SRM ATC RTT GGT GAC
    ATC CGT CAG GCT CAC TGC AAC ATC TCT CGT GCT AAA TGG
    AAC AAC ACT-3' (Seq ID No. 8)
```

wiedergegeben und entspricht dem Satz von Polypeptid-Antigenen, der durch die folgende allgemeine Sequenz

Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn  
 Thr Arg Xaa<sub>1</sub> Xaa<sub>2</sub> Ile Xaa<sub>3</sub> Xaa<sub>4</sub> Gly Pro Gly Arg Ala Xaa<sub>5</sub> Tyr Xaa<sub>6</sub>  
 Thr Xaa<sub>7</sub> Xaa<sub>8</sub> Ile Xaa<sub>9</sub> Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Arg  
 Ala Lys Trp Asn Asn Thr (Seq ID No. 9)

wiedergegeben wird, in welcher  
 Xaa<sub>1</sub> Lys oder Arg ist,  
 Xaa<sub>2</sub> Ser oder Gly ist,  
 Xaa<sub>3</sub> His, Arg, Pro, Thr oder Asn ist,  
 Xaa<sub>4</sub> Ile oder Met ist,  
 Xaa<sub>5</sub> Phe oder Ile ist,  
 Xaa<sub>6</sub> Thr oder Ala ist,  
 Xaa<sub>7</sub> Gly oder Glu ist,  
 Xaa<sub>8</sub> Glu, Asp, Arg, Lys oder Gln ist und  
 Xaa<sub>9</sub> Ile oder Val ist.

**[0075]** Als die Population der analysierten V3-Loops von HIV-1 ausgewählt wurde, so dass sie Sequenzen aus Isolaten von HIV-1 aus Uganda enthielt (Oram et al. (1991) AIDS Research and Human Retroviruses 7:605), wurde die folgende degenerierte Oligonucleotidsequenz ermittelt:

```
5' -AGCTGAACGAATCTGTTGACATCAACTGCWCCCTCCGWACAAMA
    SAYCAKAMAGRMMTSMVCMTGGCCCGGGCMRAGYTDTSYWCAC
    ACCRRAADAAYCGGCKACATCSGTCAGGCTYACTGCAACATCTC
    CGTGCTAAATGGAACAACACT-3' (Seq ID No. 10)
```

**[0076]** Dieser Satz codiert den folgenden Satz von Polypeptid-Antigenen, der durch die allgemeine Formel

Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Xaa<sub>1</sub> Arg Pro Xaa<sub>2</sub> Xaa<sub>3</sub> Xaa<sub>4</sub>

Xaa<sub>5</sub> Xaa<sub>6</sub> Xaa<sub>7</sub> Xaa<sub>8</sub> Xaa<sub>9</sub> Xaa<sub>10</sub> Xaa<sub>11</sub> Gly Pro Gly Xaa<sub>12</sub> Xaa<sub>13</sub> Xaa<sub>14</sub>

Xaa<sub>15</sub> Thr Thr Xaa<sub>16</sub> Xaa<sub>17</sub> Xaa<sub>18</sub> Gly Xaa<sub>19</sub> Ile Xaa<sub>20</sub> Gln Ala Xaa<sub>21</sub> Cys

Asn Ile Ser Arg Ala Lys Trp Asn Asn Thr (Seq ID No. 11)

wiedergegeben wird, in welcher

Xaa<sub>1</sub> Thr oder Ser ist,

Xaa<sub>2</sub> Asn oder Tyr ist,

Xaa<sub>3</sub> Asn oder Lys ist,

Xaa<sub>4</sub> Asn oder Lys ist,

Xaa<sub>5</sub> Thr oder Ile ist,

Xaa<sub>6</sub> Arg oder Ile ist,

Xaa<sub>7</sub> Lys oder Gln ist,

Xaa<sub>8</sub> Ser, Gly oder Arg ist,

Xaa<sub>9</sub> Ile, Met oder Leu ist,

Xaa<sub>10</sub> His, Asn, Arg, Ser, Pro oder Thr ist,

Xaa<sub>11</sub> Ile, Met oder Leu ist,

Xaa<sub>12</sub> Arg, Lys oder Gln ist,

Xaa<sub>13</sub> Ala oder Val ist,

Xaa<sub>14</sub> Pre, Leu, Ile, Met oder Val ist,

Xaa<sub>15</sub> Tyr, Phe, His oder Leu ist,

Xaa<sub>16</sub> Gly, Lys, Arg oder Glu ist,

Xaa<sub>17</sub> Ile, Lys oder Arg ist,

Xaa<sub>18</sub> Ile oder Thr ist,

Xaa<sub>19</sub> Asp oder Tyr ist,

Xaa<sub>20</sub> Arg oder Gly ist und

Xaa<sub>21</sub> His oder Tyr ist.

**[0077]** Um diese degenerierte Oligonucleotidsequenz zu synthetisieren, wurden die folgenden synthetischen Oligonucleotide hergestellt:

#CR11 5'-CGCGAATTCTCCATGGTTCAGCTGAACGAATCTGTTGAC-3'  
(Seq ID No. 12)

#CR22 5'-CGGACGGGWGCAGTTGATGTCAACAGATTCGTTCA  
(Seq ID No. 13)

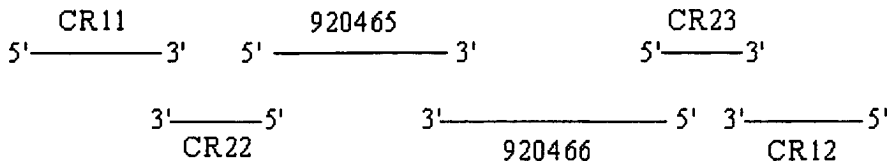
#920465 5'-ATCAACTGCWCCCGTCCGWACAAMAASAYCAKAMAGRMM  
TSMVCMTSGGCCCGG  
(Seq ID No. 14)

#920466 5'-ATGTTGCAGTRAGCCTGACSGATGTMGCCGRTTHTTTYGG  
TAGTGWRSARCTYKACCCGGGCCSAKG  
(Seq ID No. 15)

#CR23 5'-CAGGCTYACTGCAACATCTCTCGTGCTAAATGGAACA  
(Seq ID No. 16)

#CR12 5'-CGCGTCGACA GTGTTGTTC ATTAGCACG AGAG-3'  
(Seq ID No. 17)

**[0078]** Diese Oligonucleotide wurden wie folgt zusammengefügt:



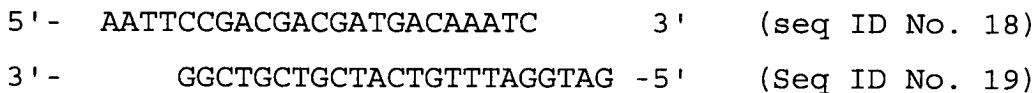
**[0079]** Von jedem der obigen Nucleotide wurde durch Auflösen der Nucleotide in sterilem Wasser eine Vorratslösung hergestellt. Aliquots der Vorratslösungen wurden mit Kinase behandelt und dann zur Hybridisierung in Klenow-Puffer und sterilem Wasser zusammen gemischt. Die Reaktionsmischung wurde auf 94°C erhitzt und langsam pro 15 Sekunden um 1°C abgekühlt. Zur Reaktionsmischung wurden sodann dGTP, dCTP, dTTP, dATP, ein Klenow-Fragment, ATP und Ligase gegeben. Die Mischung wurde über Nacht bei Raumtemperatur inkubiert. Die Mischung wurde sodann mit 95% Ethanol gefällt und das DNA-Pellet mit 70% Ethanol gewaschen, getrocknet und in 20 µl sterilem Wasser gelöst.

**[0080]** Die isolierten DNA-Sequenzen wurden unter Verwendung von CR11 und CR12 als 5'- bzw. 3'-Amplimere mittels PCR amplifiziert. Das 5'-Amplimer weist EcoR I-, Nco I- und Pvu II-Stellen auf und das 3'-Amplimer verfügt über eine Sa II-Stelle, mit der in verschiedene Expressionssysteme kloniert werden kann.

**[0081]** Die PCR-Produkte wurden mit Hilfe der Gelelektrophorese isoliert, mit "Gene clean" (Bio101) gereinigt und mit den Restriktionsenzymen EcoR I und Sa II zerschnitten. Die restringierte DNA-Bibliothek wurde sodann in ein pFLAG-Plasmid (IBI FLAG Biosystem, Katalog-Nr. IB 13000) kloniert und mit EcoR I und Sa II und dann mit alkalischer Phosphatase aus Kälberdarm behandelt. Die so hergestellte Vektor-Bibliothek codiert eine im Leseraster verlaufende Fusion des Gens für das FLAG-Peptid und der Varianten des V3-Gens. Nach Induktion mit IPTG wurde eine Bibliothek von aus dem FLAG-Peptid (Amino-Ende) und den Varianten des V3-Loops (Carboxy-Ende) zusammengesetzten Fusions-Polypeptiden hergestellt.

**[0082]** Die PCR-Produkte (Genbibliothek aus V3-Loops) können auch mit Pvu II und Sa II geschnitten und in das Expressionssystem pEZZmp18 oder pEBmp18 ligiert werden (Stahl et al. (1989) J. Immunol. Meth. 124:43), um eine Bibliothek aus Fusionsproteinen mit einem Staphylococcus-Protein und den V3-Sequenzen herzustellen. Alle drei Plasmide enthalten das pBR322 ori um die Replikation im passenden Wirtsorganismen voran zu treiben sowie F1 ori, um eine zielgerichtete Mutagenese zu erleichtern.

**[0083]** Die oben erzeugten PCR-Produkte wurden z.B. mit Nco I und Sa II geschnitten und an das doppelsträngige Oligonucleotid



ligiert, welches eine Erkennungssequenz für die Enterokinase-Spaltung (EKCR) im Leseraster mit der den V3-Loop codierenden Sequenz codiert.

**[0084]** Das erhaltene EKCRN3-Loop-Fusionsgen wurde dann unter Verwendung der durch Behandlung des EKCR/V3-Fusionsgens mit den entsprechenden Endonucleasen erhaltenen EcoR I- und Sa II-Überhänge in die EcoR I- und Sa II-Stellen des pEZZ-18 ligiert (Pharmazia Katalog-Nr. 27-4810-01). Der Vektor pEZZ-18 die Signalsequenz für das Protein A und zwei synthetische "Z"-Domänen, die auf der "B"-IgG-Bindungsdomäne des Proteins A basieren. Dieses Konstrukt bewirkt, dass "ZZ"-Fusionsproteine aus E. coli ausgeschieden werden und die Löslichkeit in wässrigen Umgebungen erhöht und die Affinitätsreinigung des Fusionsproteins auf Säulen mit IgG erleichtert wird. Somit codiert das erhaltene Fusionsgen ein ZZ/EKCR/V3-Loop-Fusionsprotein. Durch Behandlung des erhaltenen Fusionsproteins mit Enterokinase lassen sich die Sequenzen von Protein A aus V3 entfernen. Siehe Su et al. (1992) Biotechniques 13:756 sowie Forsberg et al. (1992) J. Protein Chem. 11:201.

**[0085]** Die pEZZ-18-Konstrukte wurden erzeugt und zur Transformation kompetenter Zellen eingesetzt. Die erhaltenen Fusionsproteine wurden sodann gereinigt. Kurz gesagt wurden die das ZZ/EKCR/V3-Konstrukt tragenden Zellen in 2xYT-Medium gezüchtet. Die Zellen wurden in einer späten stationären Phase des Wachstums geerntet und in einem Puffer (20 mM Na-Acetat, pH 5,5, 1 mM PMSF, 5 mM CHAPS, 10% Glycerin, 2 µg/ml Aprotinin) unter Beschallung resuspendiert. Nach der Beschallung wurde das Lysat bei 10.000 Upm in einem Beckmann JA-17-Rotor 15 Minuten lang zentrifugiert. Der Überstand wurde dann nach Vorschrift des Herstellers zur Reinigung auf eine Affinitätssäule mit IgG gegeben. Der affinitätsgereinigte V3-Loop wurde zur weiteren Reinigung einer PAGE und Elektroelution unterzogen.

**[0086]** Der Satz von Polypeptid-Antigenen kann dazu verwendet werden, mit standardisierten Immunisierungsverfahren polyklonale Seren in Kaninchen sowie eine gereinigte Mischung aus polyklonalen Antikörpern anzureichern. Die HN-Infektiosität sowie gp120/CD4-Bindungsassays können eingesetzt werden, um die Wirksamkeit des Satzes von Polypeptid-Antigenen zu testen, indem eine Immunantwort (z.B. eine Antikörper-Reaktion) gegen Varianten von HIV hervorgerufen wird.

**[0087]** Die polyklonalen Seren können auch dazu verwendet werden, künstlich einen selektiven Mutationsdruck auf das Virus auszuüben, um den evolutionären Zeitrahmen zu verkürzen. Beispielsweise lassen sich mit HIV infizierte Zellen, die in der Lage sind, so zu mutieren, dass die neue Variante der Erkennung durch polyklonale Seren zu entgeht, nachweisen. Diese Varianten werden sequenziert und in ausgewogener Weise in einer Populationsanalyse verwendet, so dass sie in einem folgenden Satz von Polypeptid-Antigenen enthalten sind.

SEQUENZPROTOKOLL

(1) ALLGEMEINE INFORMATION

- (i) ANMELDER: CREA, Roberto
  
- (ii) TITEL DER ERFINDUNG: KOMBINATORISCHE POLYPEPTID-ANTIGENE
  
- (iii) ANZAHL DER SEQUENZEN: 19
- (iv) KORRESPONDENZADRESSE:
  - (A) ADRESSE: LAHIVE & COCKFIELD
  - (B) STRASSE: 60 STATE STREET, SUITE 510
  - (C) STADT: BOSTON
  - (D) STAAT: MA
  - (E) LAND: USA
  - (F) POSTLEITZAHL (ZIP): 021=)
  
- (v) COMPUTER LESBARE FORM
  - (A) ART DES INFORMATIONSSPEICHERS: FLOPPY DISK
  - (B) COMPUTER: IBM PC kompatibel
  - (C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: ASCII-Text
  
- (vi) DATEN FÜR LAUFENDE ANMELDUNG:
  - (A) ANMELDENUMMER:
  - (B) DATUM DER EINREICHUNG: 18. JUNI 1993
  - (C) KLASSIFIZIERUNG:
  
- (vii) FRÜHERE PATENTANMELDUNGEN:
  - (A) ANMELDENUMMER: US 07/900,123
  - (B) DATUM DER EINREICHUNG: 18 JUNI 1992
  
- (viii) PATENTANWALT:
  - (A) NAME: DeConti, Giulio A.
  - (B) REGISTRIERUNGSNUMMER: 31,503
  - (C) AKTENZEICHEN: CTE-003PC
  
- (ix) INFORMATION ÜBER TELEKOMMUNIKATION:
  - (A) TELEFON: (617) 227-7400
  - (B) TELEFAX: (617) 227-5941

(2) INFORMATION ZU SEQ ID NO: 1:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 35 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro  
 1 5 10 15  
 Gly Arg Ala Leu Tyr Thr Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln  
 20 25 30  
 Ala His Cys  
 35

(2) INFORMATION ZU SEQ ID NO: 2:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 36 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(v) ART DES FRAGMENTS: intern

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Lys Ile Arg Ile Gln Arg  
 1 5 10 15  
 Gly Pro Gly Arg Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met Arg  
 20 25 30  
 Gln Ala His Cys  
 35

(2) INFORMATION ZU SEQ ID NO: 3:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 35 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(v) ART DES FRAGMENTS: intern

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Thr Ile Gly Pro  
1                   5                   10                   15  
  
Gly Arg Ala Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln  
          20                   25                   30  
  
Ala His Cys  
      35

(2) INFORMATION ZU SEQ ID NO: 4:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 34 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(v) ART DES FRAGMENTS: intern

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Lys Ile Arg Ile Gly Pro  
1                   5                   10                   15  
  
Gly Arg Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met Arg Gln Ala  
          20                   25                   30  
  
His Cys

(2) INFORMATION ZU SEQ ID NO: 5:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 35 Aminosäuren

- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(v) ART DES FRAGMENTS: intern

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 5:

Cys	Thr	Arg	Pro	Asn	Asn	Asn	Thr	Arg	Lys	Ser	Ile	His	Ile	Gly	Pro
1				5					10					15	
Gly	Arg	Ala	Phe	Tyr	Thr	Thr	Gly	Glu	Ile	Ile	Gly	Asp	Ile	Arg	Gln
		20					25					30			
Ala	His	Cys													
		35													

(2) INFORMATION ZU SEQ ID NO: 6:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 165 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRÄNGIGKEIT: einzelsträngig
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 6:

G TTCAGCTGA	ACGAATCTGT	TGACATCAAC	TGCAYCCGTC	CGARCAACAA	CACCARAARA	60
RGMATCMVCA	TSGGCCCGGG	CARAGYTWTC	YACRCTAYCS	RGSRMWTCRY	TGGTRACATC	120
CGTMAGGCTC	ACTGCAACAT	CTCTCGTGCT	AAATGGAACA	ACACT		165

(2) INFORMATION ZU SEQ ID NO: 7:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 55 Aminosäuren
  - (B) ART: Aminosäure
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid



(v) ART DES FRAGMENTS: intern

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 12
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Thr oder Ile"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 15
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Asn oder Ser"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 19
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Arg oder Lys"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 20
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Arg oder Lys"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 21
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Arg, Gly oder Ser"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 23
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Asn, Pro, Ser, Arg, Thr oder His"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 24
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Met oder Ile"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 28

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Lys oder Arg"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 29

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Val oder Ala"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 30

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile oder Phe"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 30

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist His oder Tyr"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 32

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ala oder Thr"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 33

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile oder Thr"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 34

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Arg, Asn, Glu  
oder Gly"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 35

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Gly, Arg, His,  
Gln, Asp oder Glu"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 36

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Phe oder Ile"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 37

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ala, Thr, Val  
oder Ile"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 39

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Asn oder Asp"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 42

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Lys oder Gln"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 7:

```

Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Xaa Arg Pro Xaa Asn
1           5           10           15
Asn Thr Xaa Xaa Xaa Ile Xaa Xaa Gly Pro Gly Xaa Xaa Xaa Xaa Xaa
20           25           30
Xaa Xaa Xaa Xaa Xaa Gly Xaa Ile Arg Xaa Ala His Cys Asn Ile Ser
35           40           45
Arg Ala Lys Trp Asn Asn Thr
50           55
    
```

(2) INFORMATION ZU SEQ ID NO: 8:

(i) EIGENSCHAFTEN DER SEQUENZ:

(A) LÄNGE: 165 Basenpaare

(B) ART: Nucleinsäure

(C) STRÄNGIGKEIT: einzelsträngig

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 8:

G TTCAGCTGA ACGAATCTGT TGACATCAAC TGCACCCGTC CGAACAACAA CACCCGTARA 60  
 RGCATCMVCA TSGGCCCGGC CCGTGCTWTC TACRCTACCG RGSRMATCRT TGGTGACATC 120  
 CGTCAGGCTC ACTGCAACAT CTCTCGTGCT AAATGGAACA AACT 165

## (2) INFORMATION ZU SEQ ID NO: 9:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 55 Aminosäuren
  - (B) ART: Aminosäure
  - (D) TOPOLOGIE: linear
- (ii) ART DES MOLEKÜLS: Peptid
- (v) ART DES FRAGMENTS: intern
- (ix) MERKMAL:
  - (A) NAME/SCHLÜSSEL: ortsmodifiziert
  - (B) LAGE: 20
  - (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Lys oder Arg"
- (ix) MERKMAL:
  - (A) NAME/SCHLÜSSEL: ortsmodifiziert
  - (B) LAGE: 21
  - (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ser oder Gly"
- (ix) MERKMAL:
  - (A) NAME/SCHLÜSSEL: ortsmodifiziert
  - (B) LAGE: 23
  - (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist His, Arg, Pro, Thr oder Asn"
- (ix) MERKMAL:
  - (A) NAME/SCHLÜSSEL: ortsmodifiziert
  - (B) LAGE: 24
  - (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile oder Met"
- (ix) MERKMAL:
  - (A) NAME/SCHLÜSSEL: ortsmodifiziert
  - (B) LAGE: 30
  - (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Phe oder Ile"
- (ix) MERKMAL:
  - (A) NAME/SCHLÜSSEL: ortsmodifiziert

- (B) LAGE: 32
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Thr oder Ala"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 34
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Gly oder Glu"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 35
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Glu, Asp, Arg, Lys oder Gln"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 37
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile oder Val"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 9:

```

Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn
1           5           10           15
Asn Thr Arg Xaa Xaa Ile Xaa Xaa Gly Pro Gly Arg Ala Xaa Tyr Xaa
20           25           30
Thr Xaa Xaa Ile Xaa Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser
35           40           45
Arg Ala Lys Trp Asn Asn Thr
50           55
    
```

(2) INFORMATION ZU SEQ ID NO: 10:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 162 Basenpaare
- (B) ART: Nucleinsäure
- (C) STRÄNGIGKEIT: einzelsträngig
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 10:

G TTCAGCTGA ACGAATCTGT TGACATCAAC TGCWCCCGTC CGWACAAMAA SAYCAKAMAG

RGMMTSMVCM TSGGCCCGGG CMRAGYTDTS YWCACTACCR RAADAAYCGG CKACATCSGT 120  
CAGGCTYACT GCAACATCTC TCGTGCTAAA TGGAACAACA CT 162

## (2) INFORMATION ZU SEQ ID NO: 11:

## (i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 54 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

## (ii) ART DES MOLEKÜLS: Peptid

## (v) ART DES FRAGMENTS: intern

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 12
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Thr oder Ser"

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 15
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Asn oder Tyr"

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 16
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Asn oder Lys"

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 17
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Asn oder Lys"

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 18
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Thr oder Ile"

## (ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 19
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Arg oder Ile"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 20
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Lys oder Gln"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 21
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ser, Gly oder Arg"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 22
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile, Met oder Leu"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 23
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist His, Asn, Arg, Ser, Pro oder Thr"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 24
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile, Met oder Leu"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 28
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Arg, Lys oder Gln"

(ix) MERKMAL:

- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 29

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ala oder Val"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 30

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Pre, Leu, Ile,  
Met oder Val"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 31

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Tyr, Phe, His  
oder Leu"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 34

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Gly, Lys, Arg  
oder Glu"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 35

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile, Lys oder  
Arg"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 36

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Ile oder Thr"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 38

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Asp oder Tyr"

(ix) MERKMAL:

(A) NAME/SCHLÜSSEL: ortsmodifiziert

(B) LAGE: 40

(D) WEITERE INFORMATION: /Bemerkung: "Xaa ist Arg oder Gly"

(ix) MERKMAL:



- (A) NAME/SCHLÜSSEL: ortsmodifiziert
- (B) LAGE: 43
- (D) WEITERE INFORMATION: /Bemerkung: "Xaa ist His oder Tyr"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 11:

Val	Gln	Leu	Asn	Glu	Ser	Val	Glu	Ile	Asn	Cys	Xaa	Arg	Pro	Xaa	Xaa
1				5					10					15	
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Gly	Pro	Gly	Xaa	Xaa	Xaa	Xaa	Thr
			20					25					30		
Thr	Xaa	Xaa	Xaa	Gly	Xaa	Ile	Xaa	Gln	Ala	Xaa	Cys	Asn	Ile	Ser	Arg
		35				40						45			
Ala	Lys	Trp	Asn	Asn	Thr										
	50														

(2) INFORMATION ZU SEQ ID NO: 12:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 39 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRÄNGIGKEIT: einzelsträngig
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 12:

CGCGAATTCT CCATGGTTCA GCTGAACGAA TCTGTTGAC

39

(2) INFORMATION ZU SEQ ID NO: 13:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 35 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRÄNGIGKEIT: einzelsträngig
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 13:

CGGACGGGGWG CAGTTGATGT CAACAGATTC GTTCA

35

## (2) INFORMATION ZU SEQ ID NO: 14:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 55 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRÄNGIGKEIT: einzelsträngig
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 14:

ATCAACTGCW CCCGTCCGWA CAAMAASAYC AKAMAGRGMM TSMVCMTSGG CCCGG

55

## (2) INFORMATION ZU SEQ ID NO: 15:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 69 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRÄNGIGKEIT: einzelsträngig
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 15:

ATGTTGCAGT RAGCCTGACS GATGTMGCCG RTHTTTYGG TAGTGWRSAH ARCTYKACCC

60

GGCCSAKG

69

## (2) INFORMATION ZU SEQ ID NO: 16:

- (i) EIGENSCHAFTEN DER SEQUENZ:
  - (A) LÄNGE: 37 Basenpaare
  - (B) ART: Nucleinsäure
  - (C) STRÄNGIGKEIT: einzelsträngig
  - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 16:

CAGGCTYACT GCAACATCTC TCGTGCTAAA TGGAACA

37

(2) INFORMATION ZU SEQ ID NO: 17:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 34 Basenpaare
- (B) ART: Nucleinsäure
- (C) STRÄNGIGKEIT: einzelsträngig
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 17:

CGCGTCGACA GTGTTGTTCC ATTTAGCACG AGAG

34

(2) INFORMATION ZU SEQ ID NO: 18:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 23 Basenpaare
- (B) ART: Nucleinsäure
- (C) STRÄNGIGKEIT: einzelsträngig
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 18:

AATCCGACG ACGATGACAA ATC

23

(2) INFORMATION ZU SEQ ID NO: 19:

(i) EIGENSCHAFTEN DER SEQUENZ:

- (A) LÄNGE: 23 Basenpaare
- (B) ART: Nucleinsäure

(C) STRÄNGIGKEIT: einzelsträngig

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 19:

GGCTGCTGCT ACTGTTTAGG TAG

23

**Patentansprüche**

1. Verfahren zur Herstellung eines Satzes von Polypeptid-Antigenen, die Aminosäure-Sequenzen aufweisen, die sich von einem Protein oder einem Teil davon ableiten, umfassend die Schritte:

a. Auswahl des Proteins oder des Teils davon, aufweisend eine Population von Z Varianten, dargestellt durch die Formel

$$A_1 A_2 A_3 \dots A_{n-2} A_{n-1} A_n,$$

wobei  $A_n$  eine Aminosäure ist, die an der Aminosäure-Position n des Proteins oder Teils davon vorkommt;

b. Bestimmung der Anzahl der Male  $O_n^{aa}$ , die jeder Typ von Aminosäure an jeder Aminosäure-Position n in den Z Varianten vorkommt;

c. Berechnen der Häufigkeit des Vorkommens  $(O_n^{aa}/Z)_n$  jedes Typs von Aminosäure an jeder Aminosäure-Position n in den Z Varianten; und

d. Synthetisieren des Satzes von Polypeptid-Antigenen, die Aminosäure-Sequenzen aufweisen, die im Wesentlichen durch die Formel

$$A'_1 A'_2 A'_3 \dots A'_{n-2} A'_{n-1} A'_n$$

dargestellt sind, wobei  $A'_n$  als ein Typ von Aminosäure definiert ist, der in größerer als einer ausgewählten Häufigkeit an der entsprechenden Aminosäure-Position in den Z Varianten vorkommt.

2. Verfahren nach Anspruch 1, wobei der Satz von Polypeptid-Antigenen von 8 bis 150 Aminosäuren reicht.

3. Verfahren nach Anspruch 1, wobei die ausgewählte Häufigkeit 5% beträgt.

4. Verfahren nach Anspruch 1, wobei die ausgewählte Häufigkeit 10% beträgt.

5. Verfahren nach Anspruch 1, wobei das Protein eine immunogene Komponente eines Pathogens ist.

6. Verfahren nach Anspruch 5, wobei die immunogene Komponente des Pathogens ein oder mehrere neutralisierende antigene Epitope enthält.

7. Verfahren nach Anspruch 5, wobei das Protein eine Komponente eines Virus ist.

8. Verfahren nach Anspruch 7, wobei das Protein eine Komponente von HIV-1 ist.

9. Verfahren nach Anspruch 8, wobei das Protein gp120 oder ein Teil davon ist.

10. Verfahren nach Anspruch 9, wobei der Teil eines Proteins die V3-Schleife von gp120 ist.

11. Verfahren nach Anspruch 1, wobei der Satz von Polypeptid-Antigenen generiert wird durch

i) Synthetisieren einer degenerierten Oligonukleotid-Sequenz, die eine minimale Anzahl an Nukleotid-Kombinationen an jeder Codon-Position n aufweist, wobei die Kombinationen wenigstens die Codons, die für jeden Typ von Aminosäure  $A'_1$  bis  $A'_n$  codieren, umfassen; und

ii) Expression des degenerierten Oligonukleotids in einem Expressionssystem zur Herstellung eines Satzes von Polypeptid-Antigenen:

$A'_1 A'_2 A'_3 \dots A'_{n-2} A'_{n-1} A'_n$

Es folgt kein Blatt Zeichnungen