(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 November 2002 (28.11.2002)

PCT

(10) International Publication Number
WO 02/094868 A2

(51) International Patent Classification[7]: C07K 14/31

(21) International Application Number: PCT/IB02/02637

(22) International Filing Date: 27 March 2002 (27.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0107661.1 27 March 2001 (27.03.2001) GB

(71) Applicant *(for all designated States except US)*: **CHIRON SPA** [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MASIGNANI, Vega** [IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT). **MORA, Marirosa** [IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT). **SCARSELLI, Maria** [IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT).

(74) Agents: **HALLYBONE, Huw, George** et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

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

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/094868 A2

(54) Title: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS

(57) Abstract: The invention provides proteins from *Staphylococcus aureus* including amino acid sequences and the corresponding nucleotide sequences. The proteins are useful for vaccines, immunogenic compositions, diagnostics, enzymatic studies and also as targets for antibiotics.

STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention relates to nucleic acid and proteins from the bacteria *Staphylococcus aureus*.

BACKGROUND ART

Staphylococcus aureus is a Gram-positive spherical bacterium which, on microscopic examination, appears in pairs, short chains, or bunched clusters. Some strains are capable of producing a highly heat-stable protein enterotoxin that causes food poisoning (staphyloenterotoxicosis) in humans. Of particular clinical concern are strains which are resistant to a wide range of antibiotics ('MRSA').

There is currently no effective vaccine against *S.aureus*, although a polysaccharide conjugate vaccine is currently undergoing clinical trials (StaphVAX™ from Nabi).

It is an object of the invention to provide proteins which can be used in the development of vaccines. Further objects are to provide proteins and nucleic acid which can be used in the diagnosis of *S.aureus* infection, to provide proteins and nucleic acid which can be used for the detection of *S.aureus*, to provide nucleic acid which is useful for the expression of *S.aureus* proteins, and to provide proteins which are useful targets for antibiotic research.

DISCLOSURE OF THE INVENTION

The invention provides proteins comprising the *S.aureus* amino acid sequences disclosed in the examples. These amino acid sequences are the even SEQ IDs between 2 and 5642.

It also provides proteins comprising amino acid sequences having sequence identity to the *S.aureus* amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of sequence identity is preferably greater than 50% (*e.g.* 60%, 70%, 80%, 90%, 95%, 99% or more). These proteins include homologs, orthologs, allelic variants and functional mutants. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the *S.aureus* amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more). Preferably the fragments comprise one or more epitopes from the sequence. Other preferred fragments are (a) the N-terminal signal peptides of the proteins disclosed in the examples, (b) the proteins disclosed in the examples, but without their N-terminal signal peptides, and (c) the proteins disclosed in the examples, but without their N-terminal amino acid residue..

The proteins of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from *S.aureus*, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other staphylococcal or host cell proteins). Proteins of the invention are preferably staphylococcal proteins.

Preferred proteins are those which show homology to the 'GBS*nnn*' antigens listed in Table IV of PCT/GB01/04789 *i.e.* which are inferred to be useful antigens for immunisation and/or diagnosis. These include SEQ IDs: 1078 (GBS199), 1876 (GBS177), 1946 (GBS311), 2508 (GBS312), 3724 (GBS25), 4600 (GBS90), 4826 (GBS492), and 5360 (GBS114).

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means. The antibodies may include a detectable label.

According to a further aspect, the invention provides nucleic acid comprising the *S.aureus* nucleotide sequences disclosed in the examples. These sequences are the odd SEQ IDs between 1 and 5641.

In addition, the invention provides nucleic acid comprising nucleotide sequences having sequence identity to the *S.aureus* nucleotide sequences disclosed in the examples. Identity between sequences is preferably determined by the Smith-Waterman homology algorithm as described above.

Furthermore, the invention provides nucleic acid which can hybridise to the *S.aureus* nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*e.g.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least *n* consecutive nucleotides from the *S.aureus* sequences and, depending on the particular sequence, *n* is 10 or more (*e.g.* 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more).

Nucleic acids of the invention can be used in hybridisation reactions (*e.g.* Northern or Southern blots, or in nucleic acid microarrays or 'gene chips') and amplification reactions (*e.g.* PCR, SDA, SSSR, LCR, NASBA, TMA) *etc.*

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*e.g.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*e.g.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*e.g.* single stranded, double stranded, vectors, primers, probes *etc.*). The nucleic acid is preferably in substantially isolated form.

Nucleic acid according to the invention may be labelled *e.g.* with a radioactive or fluorescent label. This is particularly useful where it is to be used as a primer or probe *e.g.* in PCR, LCR, TMA, NASBA.

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*e.g.* cloning or expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as immunogenic compositions, for instance, or as diagnostic reagents, or as vaccines.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*e.g.* as immunogenic compositions or vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to staphylococcus; (ii) a diagnostic reagent for detecting the presence of staphylococcus or of antibodies raised against staphylococcus; and/or (iii) a reagent which can raise antibodies against staphylococcus. Said staphylococcus may be any species, group or strain, but is preferably *S.aureus*.

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody of the invention.

The invention also provides a kit comprising primers (*e.g.* PCR primers) for amplifying a target sequence contained within a *Staphylococcus* (*e.g. S.aureus*) nucleic acid sequence, the kit comprising a first primer and a second primer, wherein the first primer is substantially complementary to said target sequence and the second primer is substantially complementary to a complement of said target sequence, wherein the parts of said primers which have substantial complementarity define the termini of the target sequence to be amplified. The first primer and/or the second primer may include a detectable label (*e.g.* a fluorescent label).

The invention also provides a kit comprising first and second single-stranded oligonucleotides which allow amplification of a *Staphylococcus* (*e.g. S.aureus*) template nucleic acid sequence contained in a single- or double-stranded nucleic acid (or mixture thereof), wherein: (a) the first oligonucleotide comprises a primer sequence which is substantially complementary to said template nucleic acid sequence; (b) the second oligonucleotide comprises a primer sequence which is substantially complementary to the complement of said template nucleic acid sequence; (c) the first oligonucleotide and/or the second oligonucleotide comprise(s) sequence which is not complementary to said template nucleic acid; and (d) said primer sequences define the termini of the template sequence to be amplified. The non-complementary sequence(s) of feature (c) are preferably upstream of (*i.e.* 5' to) the primer sequences. One or both of these (c) sequences may comprise a restriction site (*e.g.* EP-B-0509612) or a

promoter sequence (*e.g.* EP-B-0505012). The first oligonucleotide and/or the second oligonucleotide may include a detectable label (*e.g.* a fluorescent label).

The template sequence may be any part of a genome sequence. For example, it could be a rRNA gene or a protein-coding gene. The template sequence is preferably specific to *S.aureus*.

The invention also provides a hybrid protein represented by the formula NH_2-A-[-X-L-]$_n$-B-COOH, wherein X is an amino acid sequence of the invention as described above, L is an optional linker amino acid sequence, A is an optional N-terminal amino acid sequence, B is an optional C-terminal amino acid sequence, and n is an integer greater than 1. The value of n is between 2 and x, and the value of x is typically 3, 4, 5, 6, 7, 8, 9 or 10. Preferably n is 2, 3 or 4; it is more preferably 2 or 3; most preferably, n = 2. For each n instances, -X- may be the same or different. For each n instances of [-X-L-], linker amino acid sequence -L- may be present or absent. For instance, when n=2 the hybrid may be NH_2-X_1-L_1-X_2-L_2-COOH, NH_2-X_1-X_2-COOH, NH_2-X_1-L_1-X_2-COOH, NH_2-X_1-X_2-L_2-COOH, *etc.* Linker amino acid sequence(s) -L- will typically be short (*e.g.* 20 or fewer amino acids *i.e.* 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include short peptide sequences which facilitate cloning, poly-glycine linkers (*i.e.* Gly_n where n = 2, 3, 4, 5, 6, 7, 8, 9, 10 or more), and histidine tags (*i.e.* His_n where n = 3, 4, 5, 6, 7, 8, 9, 10 or more). Other suitable linker amino acid sequences will be apparent to those skilled in the art. -A- and -B- are optional sequences which will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include leader sequences to direct protein trafficking, or short peptide sequences which facilitate cloning or purification (*e.g.* histidine tags *i.e.* His_n where n = 3, 4, 5, 6, 7, 8, 9, 10 or more). Other suitable N-terminal and C-terminal amino acid sequences will be apparent to those skilled in the art.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell of to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridising conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting *Staphylococcus* in a biological sample is also provided, comprising the step of contacting nucleic according to the invention with the biological sample under hybridising conditions. The process may involve nucleic acid amplification (*e.g.* PCR, SDA, SSSR, NASBA, LCR, TMA *etc.*) or hybridisation (*e.g.* microarrays, blots, hybridisation with a probe in solution *etc.*). PCR detection of *S.aureus* in clinical samples (especially MRSA) has previously been reported [see *e.g.* Murakami *et al.* (1991) *J. Clin. Microbiol.*29:2240-2244; Bignardi *et al.* (1996) *J. Antimicrob. Chemother.* 37:53-63;

Kitagawa *et al.* (1996) *Ann. Surgery* 224:665-671; Vannuffel *et al.* (1998) *J. Clin. Microbiol.* 36:2366-2368; Jayaratne & Rutherford (1999) *Diagn. Microbiol. Infect. Dis.* 35:13-18].

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody of the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

The invention also provides a process for determining whether a test compound binds to a protein of the invention. If a test compound binds to a protein of the invention and this binding inhibits the life cycle of the *S.aureus* bacterium, then the test compound can be used as an antibiotic or as a lead compound for the design of antibiotics. The process will typically comprise the steps of contacting a test compound with a protein of the invention, and determining whether the test compound binds to said protein. Preferred proteins of the invention for use in these processes are enzymes (*e.g.* tRNA synthetases), membrane transporters and ribosomal proteins. Suitable test compounds include proteins, polypeptides, carbohydrates, lipids, nucleic acids (*e.g.* DNA, RNA, and modified forms thereof), as well as small organic compounds (*e.g.* MW between 200 and 2000 Da). The test compounds may be provided individually, but will typically be part of a library (*e.g.* a combinatorial library). Methods for detecting a binding interaction include NMR, filter-binding assays, gel-retardation assays, displacement assays, surface plasmon resonance, reverse two-hybrid *etc.* A compound which binds to a protein of the invention can be tested for antibiotic activity by contacting the compound with *S.aureus* and then monitoring for inhibition of growth. The invention also provides a compound identified using these methods.

The invention also provides a composition comprising a protein or the invention and one or more of the following antigens:

- a protein antigen from *Helicobacter pylori* such as VacA, CagA, NAP, HopX, HopY [*e.g.* WO98/04702] and/or urease.
- a protein antigen from *N.meningitidis* serogroup B, such as those in WO99/24578, WO99/36544, WO99/57280, WO00/22430, Tettelin *et al.* (2000) *Science* 287:1809-1815, Pizza *et al.* (2000) *Science* 287:1816-1820 and WO96/29412, with protein '287' and derivatives being particularly preferred.
- an outer-membrane vesicle (OMV) preparation from *N.meningitidis* serogroup B, such as those disclosed in WO01/52885; Bjune *et al.* (1991) *Lancet* 338(8775):1093-1096; Fukasawa *et al.* (1999) *Vaccine* 17:2951-2958; Rosenqvist *et al.* (1998) *Dev. Biol. Stand.* 92:323-333 *etc.*
- a saccharide antigen from *N.meningitidis* serogroup A, C, W135 and/or Y, such as the oligosaccharide disclosed in Costantino *et al.* (1992) *Vaccine* 10:691-698from serogroup C [see also Costantino *et al.* (1999) *Vaccine* 17:1251-1263].
- a saccharide antigen from *Streptococcus pneumoniae* [*e.g.* Watson (2000) *Pediatr Infect Dis J* 19:331-332; Rubin (2000) *Pediatr Clin North Am* 47:269-285, v; Jedrzejas (2001) *Microbiol Mol Biol Rev* 65:187-207].

- an antigen from hepatitis A virus, such as inactivated virus [*e.g.* Bell (2000) *Pediatr Infect Dis J* 19:1187-1188; Iwarson (1995) *APMIS* 103:321-326].
- an antigen from hepatitis B virus, such as the surface and/or core antigens [*e.g.* Gerlich *et al.* (1990) *Vaccine* 8 Suppl:S63-68 & 79-80].
- an antigen from hepatitis C virus [*e.g.* Hsu *et al.* (1999) *Clin Liver Dis* 3:901-915].
- an antigen from *Bordetella pertussis*, such as pertussis holotoxin (PT) and filamentous haemagglutinin (FHA) from *B.pertussis*, optionally also in combination with pertactin and/or agglutinogens 2 and 3 [*e.g.* Gustafsson *et al.* (1996) *N. Engl. J. Med.* 334:349-355; Rappuoli *et al.* (1991) *TIBTECH* 9:232-238].
- a diphtheria antigen, such as a diphtheria toxoid [*e.g.* chapter 3 of *Vaccines* (1988) eds. Plotkin & Mortimer. ISBN 0-7216-1946-0] *e.g.* the CRM_{197} mutant [*e.g.* Del Guidice *et al.* (1998) *Molecular Aspects of Medicine* 19:1-70].
- a tetanus antigen, such as a tetanus toxoid [*e.g.* chapter 4 of Plotkin & Mortimer].
- a saccharide antigen from *Haemophilus influenzae* B.
- an antigen from *N.gonorrhoeae* [*e.g.* WO99/24578, WO99/36544, WO99/57280].
- an antigen from *Chlamydia pneumoniae* [*e.g.* WO02/02606; Kalman *et al.* (1999) *Nature Genetics* 21:385-389; Read *et al.* (2000) *Nucleic Acids Res* 28:1397-406; Shirai *et al.* (2000) *J. Infect. Dis.* 181(Suppl 3):S524-S527; WO99/27105; WO00/27994; WO00/37494].
- an antigen from *S.agalactiae* [*e.g.* PCT/GB01/04789]
- an antigen from *S.pyogenes* [*e.g.* PCT/GB01/04789]
- an antigen from *Chlamydia trachomatis* [*e.g.* WO99/28475].
- an antigen from *Porphyromonas gingivalis* [*e.g.* Ross *et al.* (2001) *Vaccine* 19:4135-4142].
- polio antigen(s) [*e.g.* Sutter *et al.* (2000) *Pediatr Clin North Am* 47:287-308; Zimmerman & Spann (1999) *Am Fam Physician* 59:113-118, 125-126] such as IPV or OPV.
- rabies antigen(s) [*e.g.* Dreesen (1997) *Vaccine* 15 Suppl:S2-6] such as lyophilised inactivated virus [*e.g. MMWR Morb Mortal Wkly Rep* 1998 Jan 16;47(1):12, 19; RabAvert™].
- measles, mumps and/or rubella antigens [*e.g.* chapters 9, 10 & 11 of Plotkin & Mortimer].
- influenza antigen(s) [*e.g.* chapter 19 of Plotkin & Mortimer], such as the haemagglutinin and/or neuraminidase surface proteins.
- an antigen from *Moraxella catarrhalis* [*e.g.* McMichael (2000) *Vaccine* 19 Suppl 1:S101-107].

Where a saccharide or carbohydrate antigen is included, it is preferably conjugated to a carrier protein in order to enhance immunogenicity [*e.g.* Ramsay *et al.* (2001) *Lancet* 357(9251):195-196; Lindberg (1999) *Vaccine* 17 Suppl 2:S28-36; *Conjugate Vaccines* (eds. Cruse *et al.*) ISBN 3805549326, particularly vol. 10:48-114 *etc.*]. Preferred carrier proteins are bacterial toxins or toxoids, such as diphtheria or tetanus

toxoids. The CRM_{197} diphtheria toxoid is particularly preferred. Other suitable carrier proteins include the *N.meningitidis* outer membrane protein [*e.g.* EP-0372501], synthetic peptides [*e.g.* EP-0378881, EP-0427347], heat shock proteins [*e.g.* WO93/17712], pertussis proteins [*e.g.* WO98/58668; EP-0471177], protein D from *H.influenzae* [*e.g.* WO00/56360], toxin A or B from *C.difficile* [*e.g.* WO00/61761], *etc.* Any suitable conjugation reaction can be used, with any suitable linker where necessary.

Toxic protein antigens may be detoxified where necessary (*e.g.* detoxification of pertussis toxin by chemical and/or genetic means).

Where a diphtheria antigen is included in the composition it is preferred also to include tetanus antigen and pertussis antigens. Similarly, where a tetanus antigen is included it is preferred also to include diphtheria and pertussis antigens. Similarly, where a pertussis antigen is included it is preferred also to include diphtheria and tetanus antigens.

Antigens are preferably adsorbed to an aluminium salt.

Antigens in the composition will typically be present at a concentration of at least 1μg/ml each. In general, the concentration of any given antigen will be sufficient to elicit an immune response against that antigen.

The invention also provides compositions comprising two or more (*e.g.* 3, 4, 5) proteins of the invention.

A summary of standard techniques and procedures which may be employed to perform the invention (*e.g.* to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *eg.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and II* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a staphylococcus sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (*eg.* see US patent 5,753,235).

Expression systems

The staphylococcus nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual, 2nd ed.]*.

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus

promoter. In addition, sequences derived from non-viral genes, such as the murine metallotheionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science 236:*1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J. 4*:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci. 79*:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell 41*:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus triparite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell 41*:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci. 14*:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminater/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell 23*:175] or polyomavirus, replicate to extremely high

copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol. 9*:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol. 6*:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*eg.* Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extra-chromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) *17*:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, *42*:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli.*

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol. 69*:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, *73*:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α-interferon, Maeda et al., (1985), *Nature 315*:592; human gastrin-releasing peptide, Lebacq-Verheyden et al., (1988), *Molec. Cell. Biol. 8*:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, *82*:8404; mouse IL-3, (Miyajima et al., (1987) *Gene 58*:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, *7*:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) *3*:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays 4*:91.The DNA sequence,

when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μm in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti* , *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol. 56*:153; Wright (1986) *Nature 321*:718; Smith et al., (1983) *Mol. Cell. Biol. 3*:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol. 25*:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg.* Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, *eg.* HPLC, affinity chromatography, ion exchange chromatography, *etc.*; electrophoresis; density gradient centrifugation; solvent extraction, *etc.* As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also present in the medium, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the

references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., Nucleic Acids Research 15:2515-2535 (1987); Wirsel et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*,. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987).

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Reptr*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet. 18*:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature 198*:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res. 8*:4057; Yelverton *et al.* (1981) *Nucl. Acids Res. 9*:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature 292*:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene 25*:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci. 80*:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol. 189*:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci. 82*:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature 254*:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' and of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding

site [Sambrook *et al.* (1989) "Expression of cloned genes in Escherichia coli." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* on *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature 309*:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene 60*:197], *trpE* [Allen et al. (1987) J. Biotechnol. 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol. 135*:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology 7*:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J. 3*:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci. 82*:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various Bacillus strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in

a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various Bacillus strains integrate into the Bacillus chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol. 32*:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: Bacillus subtilis [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], Escherichia coli [Shimatake *et al.* (1981) *Nature 292*:128; Amann *et al.* (1985) *Gene 40*:183; Studier *et al.* (1986) *J. Mol. Biol. 189*:113; EP-A-0 036 776,EP-A-0 136 829 and EP-A-0 136 907], Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655]; Streptococcus lividans [Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655], Streptomyces lividans [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with $CaCl_2$ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett. 60*:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci. 85*:856; Wang *et al.* (1990) *J. Bacteriol. 172*:949, Campylobacter], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci. 69*:2110; Dower *et al.* (1988) *Nucleic Acids Res. 16*:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol. 53*:159; Taketo (1988) *Biochim. Biophys. Acta 949*:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett. 44*:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem 170*:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett. 66*:203, Staphylococcus], [Barany *et al.* (1980) *J. Bacteriol. 144*:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss

III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, Streptococcus].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast Saccharomyces cerevisiae," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably

retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene 8*:17-24], pCl/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA 81*:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol. 158*:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *eg.* Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol. 101*:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983)

Proc. Natl. Acad. Sci. USA 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol, Rev. 51*:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts:Candida albicans [Kurtz, *et al.* (1986) *Mol. Cell. Biol. 6*:142], Candida maltosa [Kunze, *et al.* (1985) *J. Basic Microbiol. 25*:141]. Hansenula polymorpha [Gleeson, *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302], Kluyveromyces fragilis [Das, *et al.* (1984) *J. Bacteriol. 158*:1165], Kluyveromyces lactis [De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:737; Van den Berg *et al.* (1990) *Bio/Technology 8*:135], Pichia guillerimondii [Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141], Pichia pastoris [Cregg, *et al.* (1985) *Mol. Cell. Biol. 5*:3376; US Patent Nos. 4,837,148 and 4,929,555], Saccharomyces cerevisiae [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*:1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163], Schizosaccharomyces pombe [Beach and Nurse (1981) *Nature 300*:706], and Yarrowia lipolytica [Davidow, *et al.* (1985) *Curr. Genet. 10*:380471 Gaillardin, *et al.* (1985) *Curr. Genet. 10*:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol. 6*:142; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; Candida]; [Gleeson *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol. 158*:1165; De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:1165; Van den Berg *et al.* (1990) *Bio/Technology 8*:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol. 5*:3376; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; US Patent Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*;1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163 Saccharomyces]; [Beach and Nurse (1981) *Nature 300*:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet. 10*:39; Gaillardin *et al.* (1985) *Curr. Genet. 10*:49; Yarrowia].

<u>*Antibodies*</u>

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying staphylococcus proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 μg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (*eg.* 1,000*g* for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (*eg.* hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (*eg.* in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I, or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (*eg.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*eg.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/ nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be used [*eg.* Robinson & Torres (1997) *Seminars in Immunol* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses *eg.* MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the

retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*eg.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC Nol VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in

US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941 and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC with accession numbers VR-977 and VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995,WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, Nature 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus,

for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Triniti virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A.Polypeptides

One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B.Hormones, Vitamins, *etc.*

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, *etc.*

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethlylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphoshatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammelar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77); Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348); Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of

these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C & E, over time these lipoproteins lose A and acquire C & E. VLDL comprises A, B, C & E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) Annu Rev. Biochem 54:699; Law (1986) Adv. Exp Med. Biol. 151:162; Chen (1986) J Biol Chem 261:12918; Kane (1980) Proc Natl Acad Sci USA 77:2465; and Utermann (1984) Hum Genet 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Techniologies, Inc., Stoughton, MA, USA. Further description of lipoproteins can be found in WO98/06437..

F.Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, *etc.*

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic aid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and purtrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Staphylococcus antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-staphylococcus antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to staphylococcus proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1μg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes

can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (Tm) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$Tm = 81 + 16.6(\log_{10}Ci) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where Ci is the salt concentration (monovalent ions) and *n* is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the staphylococcus nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native staphylococcus sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the staphylococcus sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target

nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional staphylococcus sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a staphylococcus sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a staphylococcus sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions (*e.g.* temperature, salt condition *etc.*). For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acid. The assay is described in Mullis *et al.* [*Meth. Enzymol.* (1987) 155:335-350] & US patents 4,683,195 & 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired staphylococcus sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the staphylococcus sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

MODES FOR CARRYING OUT THE INVENTION

2821 nucleic acid coding sequences were identified in *S.aureus* (strain NCTC 8325), along with their inferred translation products. The nucleic acid sequences are given in the sequence listing with odd numbers (SEQ IDs 1, 3, 5, 7, ... , 5639, 5641). Each nucleic acid sequence is followed by its inferred

translation product (SEQ IDs 2, 4, 6, 8, ... , 5640, 5642). Inferred functions are given in field <223> of the sequence listing.

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *i.e.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant proteins can also be conveniently used to prepare antibodies *e.g.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*e.g.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, 892, 894, 896, 898, 900, 902, 904, 906, 908, 910, 912, 914, 916, 918, 920, 922, 924, 926, 928, 930, 932, 934, 936, 938, 940, 942, 944, 946, 948, 950, 952, 954, 956, 958, 960, 962, 964, 966, 968, 970, 972, 974, 976, 978, 980, 982, 984, 986, 988, 990, 992, 994, 996, 998, 1000, 1002, 1004, 1006, 1008, 1010, 1012, 1014, 1016, 1018, 1020, 1022, 1024, 1026, 1028, 1030, 1032, 1034, 1036, 1038, 1040, 1042, 1044, 1046, 1048, 1050, 1052, 1054, 1056, 1058, 1060, 1062, 1064, 1066, 1068, 1070, 1072, 1074, 1076, 1078, 1080, 1082, 1084, 1086, 1088, 1090, 1092, 1094, 1096, 1098, 1100, 1102, 1104, 1106, 1108, 1110, 1112, 1114, 1116, 1118, 1120, 1122, 1124, 1126, 1128, 1130, 1132, 1134, 1136, 1138, 1140, 1142, 1144, 1146, 1148, 1150, 1152, 1154, 1156, 1158, 1160, 1162, 1164, 1166, 1168, 1170, 1172, 1174, 1176, 1178, 1180, 1182, 1184, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1222, 1224, 1226, 1228, 1230, 1232, 1234, 1236, 1238, 1240, 1242, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, 1274, 1276, 1278, 1280, 1282, 1284, 1286, 1288, 1290, 1292, 1294, 1296, 1298, 1300, 1302, 1304, 1306, 1308, 1310, 1312, 1314, 1316, 1318, 1320, 1322, 1324, 1326, 1328, 1330, 1332, 1334, 1336, 1338, 1340, 1342, 1344, 1346, 1348, 1350, 1352, 1354, 1356, 1358, 1360, 1362, 1364, 1366, 1368, 1370, 1372, 1374, 1376, 1378, 1380, 1382, 1384, 1386, 1388, 1390, 1392, 1394, 1396, 1398, 1400, 1402, 1404, 1406, 1408, 1410, 1412, 1414, 1416, 1418, 1420, 1422, 1424, 1426, 1428, 1430, 1432, 1434, 1436, 1438, 1440, 1442, 1444, 1446, 1448, 1450, 1452, 1454, 1456, 1458, 1460, 1462, 1464, 1466, 1468, 1470, 1472, 1474, 1476, 1478, 1480, 1482, 1484, 1486, 1488, 1490, 1492, 1494, 1496, 1498, 1500, 1502, 1504, 1506, 1508, 1510, 1512, 1514, 1516, 1518, 1520, 1522, 1524, 1526, 1528, 1530, 1532, 1534, 1536, 1538, 1540, 1542, 1544, 1546, 1548, 1550, 1552, 1554, 1556, 1558, 1560, 1562, 1564, 1566, 1568, 1570, 1572, 1574, 1576, 1578, 1580, 1582, 1584, 1586, 1588, 1590, 1592, 1594, 1596, 1598, 1600, 1602, 1604, 1606, 1608, 1610, 1612, 1614, 1616, 1618, 1620, 1622, 1624, 1626, 1628, 1630, 1632, 1634, 1636, 1638, 1640, 1642, 1644, 1646, 1648, 1650, 1652, 1654, 1656, 1658, 1660, 1662, 1664, 1666, 1668, 1670, 1672, 1674, 1676, 1678, 1680, 1682, 1684, 1686, 1688, 1690, 1692, 1694, 1696, 1698, 1700, 1702, 1704, 1706, 1708, 1710, 1712, 1714, 1716, 1718, 1720, 1722, 1724, 1726, 1728, 1730, 1732, 1734, 1736, 1738, 1740, 1742, 1744, 1746, 1748, 1750, 1752, 1754, 1756, 1758, 1760, 1762, 1764, 1766, 1768, 1770, 1772, 1774, 1776, 1778, 1780, 1782, 1784, 1786, 1788, 1790, 1792, 1794, 1796, 1798, 1800, 1802, 1804, 1806, 1808, 1810, 1812, 1814, 1816, 1818, 1820, 1822, 1824, 1826, 1828, 1830, 1832, 1834, 1836, 1838, 1840, 1842, 1844, 1846, 1848, 1850, 1852, 1854, 1856, 1858, 1860, 1862, 1864, 1866, 1868, 1870, 1872, 1874, 1876, 1878, 1880, 1882, 1884, 1886, 1888, 1890, 1892, 1894, 1896, 1898, 1900, 1902, 1904, 1906, 1908,

1910, 1912, 1914, 1916, 1918, 1920, 1922, 1924, 1926, 1928, 1930, 1932, 1934, 1936, 1938, 1940, 1942, 1944, 1946, 1948, 1950, 1952, 1954, 1956, 1958, 1960, 1962, 1964, 1966, 1968, 1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024, 2026, 2028, 2030, 2032, 2034, 2036, 2038, 2040, 2042, 2044, 2046, 2048, 2050, 2052, 2054, 2056, 2058, 2060, 2062, 2064, 2066, 2068, 2070, 2072, 2074, 2076, 2078, 2080, 2082, 2084, 2086, 2088, 2090, 2092, 2094, 2096, 2098, 2100, 2102, 2104, 2106, 2108, 2110, 2112, 2114, 2116, 2118, 2120, 2122, 2124, 2126, 2128, 2130, 2132, 2134, 2136, 2138, 2140, 2142, 2144, 2146, 2148, 2150, 2152, 2154, 2156, 2158, 2160, 2162, 2164, 2166, 2168, 2170, 2172, 2174, 2176, 2178, 2180, 2182, 2184, 2186, 2188, 2190, 2192, 2194, 2196, 2198, 2200, 2202, 2204, 2206, 2208, 2210, 2212, 2214, 2216, 2218, 2220, 2222, 2224, 2226, 2228, 2230, 2232, 2234, 2236, 2238, 2240, 2242, 2244, 2246, 2248, 2250, 2252, 2254, 2256, 2258, 2260, 2262, 2264, 2266, 2268, 2270, 2272, 2274, 2276, 2278, 2280, 2282, 2284, 2286, 2288, 2290, 2292, 2294, 2296, 2298, 2300, 2302, 2304, 2306, 2308, 2310, 2312, 2314, 2316, 2318, 2320, 2322, 2324, 2326, 2328, 2330, 2332, 2334, 2336, 2338, 2340, 2342, 2344, 2346, 2348, 2350, 2352, 2354, 2356, 2358, 2360, 2362, 2364, 2366, 2368, 2370, 2372, 2374, 2376, 2378, 2380, 2382, 2384, 2386, 2388, 2390, 2392, 2394, 2396, 2398, 2400, 2402, 2404, 2406, 2408, 2410, 2412, 2414, 2416, 2418, 2420, 2422, 2424, 2426, 2428, 2430, 2432, 2434, 2436, 2438, 2440, 2442, 2444, 2446, 2448, 2450, 2452, 2454, 2456, 2458, 2460, 2462, 2464, 2466, 2468, 2470, 2472, 2474, 2476, 2478, 2480, 2482, 2484, 2486, 2488, 2490, 2492, 2494, 2496, 2498, 2500, 2502, 2504, 2506, 2508, 2510, 2512, 2514, 2516, 2518, 2520, 2522, 2524, 2526, 2528, 2530, 2532, 2534, 2536, 2538, 2540, 2542, 2544, 2546, 2548, 2550, 2552, 2554, 2556, 2558, 2560, 2562, 2564, 2566, 2568, 2570, 2572, 2574, 2576, 2578, 2580, 2582, 2584, 2586, 2588, 2590, 2592, 2594, 2596, 2598, 2600, 2602, 2604, 2606, 2608, 2610, 2612, 2614, 2616, 2618, 2620, 2622, 2624, 2626, 2628, 2630, 2632, 2634, 2636, 2638, 2640, 2642, 2644, 2646, 2648, 2650, 2652, 2654, 2656, 2658, 2660, 2662, 2664, 2666, 2668, 2670, 2672, 2674, 2676, 2678, 2680, 2682, 2684, 2686, 2688, 2690, 2692, 2694, 2696, 2698, 2700, 2702, 2704, 2706, 2708, 2710, 2712, 2714, 2716, 2718, 2720, 2722, 2724, 2726, 2728, 2730, 2732, 2734, 2736, 2738, 2740, 2742, 2744, 2746, 2748, 2750, 2752, 2754, 2756, 2758, 2760, 2762, 2764, 2766, 2768, 2770, 2772, 2774, 2776, 2778, 2780, 2782, 2784, 2786, 2788, 2790, 2792, 2794, 2796, 2798, 2800, 2802, 2804, 2806, 2808, 2810, 2812, 2814, 2816, 2818, 2820, 2822, 2824, 2826, 2828, 2830, 2832, 2834, 2836, 2838, 2840, 2842, 2844, 2846, 2848, 2850, 2852, 2854, 2856, 2858, 2860, 2862, 2864, 2866, 2868, 2870, 2872, 2874, 2876, 2878, 2880, 2882, 2884, 2886, 2888, 2890, 2892, 2894, 2896, 2898, 2900, 2902, 2904, 2906, 2908, 2910, 2912, 2914, 2916, 2918, 2920, 2922, 2924, 2926, 2928, 2930, 2932, 2934, 2936, 2938, 2940, 2942, 2944, 2946, 2948, 2950, 2952, 2954, 2956, 2958, 2960, 2962, 2964, 2966, 2968, 2970, 2972, 2974, 2976, 2978, 2980, 2982, 2984, 2986, 2988, 2990, 2992, 2994, 2996, 2998, 3000, 3002, 3004, 3006, 3008, 3010, 3012, 3014, 3016, 3018, 3020, 3022, 3024, 3026, 3028, 3030, 3032, 3034, 3036, 3038, 3040, 3042, 3044, 3046, 3048, 3050, 3052, 3054, 3056, 3058, 3060, 3062, 3064, 3066, 3068, 3070, 3072, 3074, 3076, 3078, 3080, 3082, 3084, 3086, 3088, 3090, 3092, 3094, 3096, 3098, 3100, 3102, 3104, 3106, 3108, 3110, 3112, 3114, 3116, 3118, 3120, 3122, 3124, 3126, 3128, 3130, 3132, 3134, 3136, 3138, 3140, 3142, 3144, 3146, 3148, 3150, 3152, 3154, 3156, 3158, 3160, 3162, 3164, 3166, 3168, 3170, 3172, 3174, 3176, 3178, 3180, 3182, 3184, 3186, 3188, 3190, 3192, 3194, 3196, 3198, 3200, 3202, 3204, 3206, 3208, 3210, 3212, 3214, 3216, 3218, 3220, 3222, 3224, 3226, 3228, 3230, 3232, 3234, 3236, 3238, 3240, 3242, 3244, 3246, 3248, 3250, 3252, 3254, 3256, 3258, 3260, 3262, 3264, 3266, 3268, 3270, 3272, 3274, 3276, 3278, 3280, 3282, 3284, 3286, 3288, 3290, 3292, 3294, 3296, 3298, 3300, 3302, 3304, 3306, 3308, 3310, 3312, 3314, 3316, 3318, 3320, 3322, 3324, 3326, 3328, 3330, 3332, 3334, 3336, 3338, 3340, 3342, 3344, 3346, 3348, 3350, 3352, 3354, 3356, 3358, 3360, 3362, 3364, 3366, 3368, 3370, 3372, 3374, 3376, 3378, 3380, 3382, 3384, 3386, 3388, 3390, 3392, 3394, 3396, 3398, 3400, 3402, 3404, 3406, 3408, 3410, 3412, 3414, 3416, 3418, 3420, 3422, 3424, 3426, 3428, 3430, 3432, 3434, 3436, 3438, 3440, 3442, 3444, 3446, 3448, 3450, 3452, 3454, 3456, 3458, 3460, 3462, 3464, 3466, 3468, 3470, 3472, 3474, 3476, 3478, 3480, 3482, 3484, 3486, 3488, 3490, 3492, 3494, 3496, 3498, 3500, 3502, 3504, 3506, 3508, 3510, 3512, 3514, 3516, 3518, 3520, 3522, 3524, 3526, 3528, 3530, 3532, 3534, 3536, 3538, 3540, 3542, 3544, 3546, 3548, 3550, 3552, 3554, 3556, 3558, 3560, 3562, 3564, 3566, 3568, 3570, 3572, 3574, 3576, 3578, 3580, 3582, 3584, 3586, 3588, 3590, 3592, 3594, 3596, 3598, 3600, 3602, 3604, 3606, 3608, 3610, 3612, 3614, 3616, 3618, 3620, 3622, 3624, 3626, 3628, 3630, 3632, 3634, 3636, 3638, 3640, 3642, 3644, 3646, 3648, 3650, 3652, 3654, 3656, 3658, 3660, 3662, 3664, 3666, 3668,

3670, 3672, 3674, 3676, 3678, 3680, 3682, 3684, 3686, 3688, 3690, 3692, 3694, 3696, 3698, 3700, 3702, 3704, 3706, 3708, 3710, 3712, 3714, 3716, 3718, 3720, 3722, 3724, 3726, 3728, 3730, 3732, 3734, 3736, 3738, 3740, 3742, 3744, 3746, 3748, 3750, 3752, 3754, 3756, 3758, 3760, 3762, 3764, 3766, 3768, 3770, 3772, 3774, 3776, 3778, 3780, 3782, 3784, 3786, 3788, 3790, 3792, 3794, 3796, 3798, 3800, 3802, 3804, 3806, 3808, 3810, 3812, 3814, 3816, 3818, 3820, 3822, 3824, 3826, 3828, 3830, 3832, 3834, 3836, 3838, 3840, 3842, 3844, 3846, 3848, 3850, 3852, 3854, 3856, 3858, 3860, 3862, 3864, 3866, 3868, 3870, 3872, 3874, 3876, 3878, 3880, 3882, 3884, 3886, 3888, 3890, 3892, 3894, 3896, 3898, 3900, 3902, 3904, 3906, 3908, 3910, 3912, 3914, 3916, 3918, 3920, 3922, 3924, 3926, 3928, 3930, 3932, 3934, 3936, 3938, 3940, 3942, 3944, 3946, 3948, 3950, 3952, 3954, 3956, 3958, 3960, 3962, 3964, 3966, 3968, 3970, 3972, 3974, 3976, 3978, 3980, 3982, 3984, 3986, 3988, 3990, 3992, 3994, 3996, 3998, 4000, 4002, 4004, 4006, 4008, 4010, 4012, 4014, 4016, 4018, 4020, 4022, 4024, 4026, 4028, 4030, 4032, 4034, 4036, 4038, 4040, 4042, 4044, 4046, 4048, 4050, 4052, 4054, 4056, 4058, 4060, 4062, 4064, 4066, 4068, 4070, 4072, 4074, 4076, 4078, 4080, 4082, 4084, 4086, 4088, 4090, 4092, 4094, 4096, 4098, 4100, 4102, 4104, 4106, 4108, 4110, 4112, 4114, 4116, 4118, 4120, 4122, 4124, 4126, 4128, 4130, 4132, 4134, 4136, 4138, 4140, 4142, 4144, 4146, 4148, 4150, 4152, 4154, 4156, 4158, 4160, 4162, 4164, 4166, 4168, 4170, 4172, 4174, 4176, 4178, 4180, 4182, 4184, 4186, 4188, 4190, 4192, 4194, 4196, 4198, 4200, 4202, 4204, 4206, 4208, 4210, 4212, 4214, 4216, 4218, 4220, 4222, 4224, 4226, 4228, 4230, 4232, 4234, 4236, 4238, 4240, 4242, 4244, 4246, 4248, 4250, 4252, 4254, 4256, 4258, 4260, 4262, 4264, 4266, 4268, 4270, 4272, 4274, 4276, 4278, 4280, 4282, 4284, 4286, 4288, 4290, 4292, 4294, 4296, 4298, 4300, 4302, 4304, 4306, 4308, 4310, 4312, 4314, 4316, 4318, 4320, 4322, 4324, 4326, 4328, 4330, 4332, 4334, 4336, 4338, 4340, 4342, 4344, 4346, 4348, 4350, 4352, 4354, 4356, 4358, 4360, 4362, 4364, 4366, 4368, 4370, 4372, 4374, 4376, 4378, 4380, 4382, 4384, 4386, 4388, 4390, 4392, 4394, 4396, 4398, 4400, 4402, 4404, 4406, 4408, 4410, 4412, 4414, 4416, 4418, 4420, 4422, 4424, 4426, 4428, 4430, 4432, 4434, 4436, 4438, 4440, 4442, 4444, 4446, 4448, 4450, 4452, 4454, 4456, 4458, 4460, 4462, 4464, 4466, 4468, 4470, 4472, 4474, 4476, 4478, 4480, 4482, 4484, 4486, 4488, 4490, 4492, 4494, 4496, 4498, 4500, 4502, 4504, 4506, 4508, 4510, 4512, 4514, 4516, 4518, 4520, 4522, 4524, 4526, 4528, 4530, 4532, 4534, 4536, 4538, 4540, 4542, 4544, 4546, 4548, 4550, 4552, 4554, 4556, 4558, 4560, 4562, 4564, 4566, 4568, 4570, 4572, 4574, 4576, 4578, 4580, 4582, 4584, 4586, 4588, 4590, 4592, 4594, 4596, 4598, 4600, 4602, 4604, 4606, 4608, 4610, 4612, 4614, 4616, 4618, 4620, 4622, 4624, 4626, 4628, 4630, 4632, 4634, 4636, 4638, 4640, 4642, 4644, 4646, 4648, 4650, 4652, 4654, 4656, 4658, 4660, 4662, 4664, 4666, 4668, 4670, 4672, 4674, 4676, 4678, 4680, 4682, 4684, 4686, 4688, 4690, 4692, 4694, 4696, 4698, 4700, 4702, 4704, 4706, 4708, 4710, 4712, 4714, 4716, 4718, 4720, 4722, 4724, 4726, 4728, 4730, 4732, 4734, 4736, 4738, 4740, 4742, 4744, 4746, 4748, 4750, 4752, 4754, 4756, 4758, 4760, 4762, 4764, 4766, 4768, 4770, 4772, 4774, 4776, 4778, 4780, 4782, 4784, 4786, 4788, 4790, 4792, 4794, 4796, 4798, 4800, 4802, 4804, 4806, 4808, 4810, 4812, 4814, 4816, 4818, 4820, 4822, 4824, 4826, 4828, 4830, 4832, 4834, 4836, 4838, 4840, 4842, 4844, 4846, 4848, 4850, 4852, 4854, 4856, 4858, 4860, 4862, 4864, 4866, 4868, 4870, 4872, 4874, 4876, 4878, 4880, 4882, 4884, 4886, 4888, 4890, 4892, 4894, 4896, 4898, 4900, 4902, 4904, 4906, 4908, 4910, 4912, 4914, 4916, 4918, 4920, 4922, 4924, 4926, 4928, 4930, 4932, 4934, 4936, 4938, 4940, 4942, 4944, 4946, 4948, 4950, 4952, 4954, 4956, 4958, 4960, 4962, 4964, 4966, 4968, 4970, 4972, 4974, 4976, 4978, 4980, 4982, 4984, 4986, 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5002, 5004, 5006, 5008, 5010, 5012, 5014, 5016, 5018, 5020, 5022, 5024, 5026, 5028, 5030, 5032, 5034, 5036, 5038, 5040, 5042, 5044, 5046, 5048, 5050, 5052, 5054, 5056, 5058, 5060, 5062, 5064, 5066, 5068, 5070, 5072, 5074, 5076, 5078, 5080, 5082, 5084, 5086, 5088, 5090, 5092, 5094, 5096, 5098, 5100, 5102, 5104, 5106, 5108, 5110, 5112, 5114, 5116, 5118, 5120, 5122, 5124, 5126, 5128, 5130, 5132, 5134, 5136, 5138, 5140, 5142, 5144, 5146, 5148, 5150, 5152, 5154, 5156, 5158, 5160, 5162, 5164, 5166, 5168, 5170, 5172, 5174, 5176, 5178, 5180, 5182, 5184, 5186, 5188, 5190, 5192, 5194, 5196, 5198, 5200, 5202, 5204, 5206, 5208, 5210, 5212, 5214, 5216, 5218, 5220, 5222, 5224, 5226, 5228, 5230, 5232, 5234, 5236, 5238, 5240, 5242, 5244, 5246, 5248, 5250, 5252, 5254, 5256, 5258, 5260, 5262, 5264, 5266, 5268, 5270, 5272, 5274, 5276, 5278, 5280, 5282, 5284, 5286, 5288, 5290, 5292, 5294, 5296, 5298, 5300, 5302, 5304, 5306, 5308, 5310, 5312, 5314, 5316, 5318, 5320, 5322, 5324, 5326, 5328, 5330, 5332, 5334, 5336, 5338, 5340, 5342, 5344, 5346, 5348, 5350, 5352, 5354, 5356, 5358, 5360, 5362, 5364, 5366, 5368, 5370, 5372, 5374, 5376, 5378, 5380, 5382, 5384, 5386, 5388, 5390, 5392, 5394, 5396, 5398, 5400, 5402, 5404, 5406, 5408, 5410, 5412, 5414, 5416, 5418, 5420, 5422, 5424, 5426, 5428,

5430, 5432, 5434, 5436, 5438, 5440, 5442, 5444, 5446, 5448, 5450, 5452, 5454, 5456, 5458, 5460, 5462, 5464, 5466, 5468, 5470, 5472, 5474, 5476, 5478, 5480, 5482, 5484, 5486, 5488, 5490, 5492, 5494, 5496, 5498, 5500, 5502, 5504, 5506, 5508, 5510, 5512, 5514, 5516, 5518, 5520, 5522, 5524, 5526, 5528, 5530, 5532, 5534, 5536, 5538, 5540, 5542, 5544, 5546, 5548, 5550, 5552, 5554, 5556, 5558, 5560, 5562, 5564, 5566, 5568, 5570, 5572, 5574, 5576, 5578, 5580, 5582, 5584, 5586, 5588, 5590, 5592, 5594, 5596, 5598, 5600, 5602, 5604, 5606, 5608, 5610, 5612, 5614, 5616, 5618, 5620, 5622, 5624, 5626, 5628, 5630, 5632, 5634, 5636, 5638, 5640, 5642.

2. A protein having 50% or greater sequence identity to a protein according to claim 1.

3. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, 892, 894, 896, 898, 900, 902, 904, 906, 908, 910, 912, 914, 916, 918, 920, 922, 924, 926, 928, 930, 932, 934, 936, 938, 940, 942, 944, 946, 948, 950, 952, 954, 956, 958, 960, 962, 964, 966, 968, 970, 972, 974, 976, 978, 980, 982, 984, 986, 988, 990, 992, 994, 996, 998, 1000, 1002, 1004, 1006, 1008, 1010, 1012, 1014, 1016, 1018, 1020, 1022, 1024, 1026, 1028, 1030, 1032, 1034, 1036, 1038, 1040, 1042, 1044, 1046, 1048, 1050, 1052, 1054, 1056, 1058, 1060, 1062, 1064, 1066, 1068, 1070, 1072, 1074, 1076, 1078, 1080, 1082, 1084, 1086, 1088, 1090, 1092, 1094, 1096, 1098, 1100, 1102, 1104, 1106, 1108, 1110, 1112, 1114, 1116, 1118, 1120, 1122, 1124, 1126, 1128, 1130, 1132, 1134, 1136, 1138, 1140, 1142, 1144, 1146, 1148, 1150, 1152, 1154, 1156, 1158, 1160, 1162, 1164, 1166, 1168, 1170, 1172, 1174, 1176, 1178, 1180, 1182, 1184, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1222, 1224, 1226, 1228, 1230, 1232, 1234, 1236, 1238, 1240, 1242, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, 1274, 1276, 1278, 1280, 1282, 1284, 1286, 1288, 1290, 1292, 1294, 1296, 1298, 1300, 1302, 1304, 1306, 1308, 1310, 1312, 1314, 1316, 1318, 1320, 1322, 1324, 1326, 1328, 1330, 1332, 1334, 1336, 1338, 1340, 1342, 1344, 1346, 1348, 1350, 1352, 1354, 1356, 1358, 1360, 1362, 1364, 1366, 1368, 1370, 1372, 1374, 1376, 1378, 1380, 1382, 1384, 1386, 1388, 1390, 1392, 1394, 1396, 1398, 1400, 1402, 1404, 1406, 1408, 1410, 1412, 1414, 1416, 1418, 1420, 1422, 1424, 1426, 1428, 1430, 1432, 1434, 1436, 1438, 1440, 1442, 1444, 1446, 1448, 1450, 1452, 1454, 1456, 1458, 1460, 1462, 1464, 1466, 1468, 1470, 1472, 1474, 1476, 1478, 1480, 1482, 1484, 1486, 1488, 1490, 1492, 1494, 1496, 1498, 1500, 1502, 1504, 1506, 1508, 1510, 1512, 1514, 1516, 1518, 1520, 1522, 1524, 1526, 1528, 1530, 1532, 1534, 1536, 1538, 1540, 1542, 1544, 1546, 1548, 1550, 1552, 1554, 1556, 1558, 1560, 1562, 1564, 1566, 1568, 1570, 1572, 1574, 1576, 1578, 1580, 1582, 1584, 1586, 1588, 1590, 1592, 1594, 1596, 1598, 1600, 1602, 1604, 1606, 1608, 1610, 1612, 1614, 1616, 1618, 1620, 1622,

1624, 1626, 1628, 1630, 1632, 1634, 1636, 1638, 1640, 1642, 1644, 1646, 1648, 1650, 1652, 1654, 1656, 1658, 1660, 1662, 1664, 1666, 1668, 1670, 1672, 1674, 1676, 1678, 1680, 1682, 1684, 1686, 1688, 1690, 1692, 1694, 1696, 1698, 1700, 1702, 1704, 1706, 1708, 1710, 1712, 1714, 1716, 1718, 1720, 1722, 1724, 1726, 1728, 1730, 1732, 1734, 1736, 1738, 1740, 1742, 1744, 1746, 1748, 1750, 1752, 1754, 1756, 1758, 1760, 1762, 1764, 1766, 1768, 1770, 1772, 1774, 1776, 1778, 1780, 1782, 1784, 1786, 1788, 1790, 1792, 1794, 1796, 1798, 1800, 1802, 1804, 1806, 1808, 1810, 1812, 1814, 1816, 1818, 1820, 1822, 1824, 1826, 1828, 1830, 1832, 1834, 1836, 1838, 1840, 1842, 1844, 1846, 1848, 1850, 1852, 1854, 1856, 1858, 1860, 1862, 1864, 1866, 1868, 1870, 1872, 1874, 1876, 1878, 1880, 1882, 1884, 1886, 1888, 1890, 1892, 1894, 1896, 1898, 1900, 1902, 1904, 1906, 1908, 1910, 1912, 1914, 1916, 1918, 1920, 1922, 1924, 1926, 1928, 1930, 1932, 1934, 1936, 1938, 1940, 1942, 1944, 1946, 1948, 1950, 1952, 1954, 1956, 1958, 1960, 1962, 1964, 1966, 1968, 1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024, 2026, 2028, 2030, 2032, 2034, 2036, 2038, 2040, 2042, 2044, 2046, 2048, 2050, 2052, 2054, 2056, 2058, 2060, 2062, 2064, 2066, 2068, 2070, 2072, 2074, 2076, 2078, 2080, 2082, 2084, 2086, 2088, 2090, 2092, 2094, 2096, 2098, 2100, 2102, 2104, 2106, 2108, 2110, 2112, 2114, 2116, 2118, 2120, 2122, 2124, 2126, 2128, 2130, 2132, 2134, 2136, 2138, 2140, 2142, 2144, 2146, 2148, 2150, 2152, 2154, 2156, 2158, 2160, 2162, 2164, 2166, 2168, 2170, 2172, 2174, 2176, 2178, 2180, 2182, 2184, 2186, 2188, 2190, 2192, 2194, 2196, 2198, 2200, 2202, 2204, 2206, 2208, 2210, 2212, 2214, 2216, 2218, 2220, 2222, 2224, 2226, 2228, 2230, 2232, 2234, 2236, 2238, 2240, 2242, 2244, 2246, 2248, 2250, 2252, 2254, 2256, 2258, 2260, 2262, 2264, 2266, 2268, 2270, 2272, 2274, 2276, 2278, 2280, 2282, 2284, 2286, 2288, 2290, 2292, 2294, 2296, 2298, 2300, 2302, 2304, 2306, 2308, 2310, 2312, 2314, 2316, 2318, 2320, 2322, 2324, 2326, 2328, 2330, 2332, 2334, 2336, 2338, 2340, 2342, 2344, 2346, 2348, 2350, 2352, 2354, 2356, 2358, 2360, 2362, 2364, 2366, 2368, 2370, 2372, 2374, 2376, 2378, 2380, 2382, 2384, 2386, 2388, 2390, 2392, 2394, 2396, 2398, 2400, 2402, 2404, 2406, 2408, 2410, 2412, 2414, 2416, 2418, 2420, 2422, 2424, 2426, 2428, 2430, 2432, 2434, 2436, 2438, 2440, 2442, 2444, 2446, 2448, 2450, 2452, 2454, 2456, 2458, 2460, 2462, 2464, 2466, 2468, 2470, 2472, 2474, 2476, 2478, 2480, 2482, 2484, 2486, 2488, 2490, 2492, 2494, 2496, 2498, 2500, 2502, 2504, 2506, 2508, 2510, 2512, 2514, 2516, 2518, 2520, 2522, 2524, 2526, 2528, 2530, 2532, 2534, 2536, 2538, 2540, 2542, 2544, 2546, 2548, 2550, 2552, 2554, 2556, 2558, 2560, 2562, 2564, 2566, 2568, 2570, 2572, 2574, 2576, 2578, 2580, 2582, 2584, 2586, 2588, 2590, 2592, 2594, 2596, 2598, 2600, 2602, 2604, 2606, 2608, 2610, 2612, 2614, 2616, 2618, 2620, 2622, 2624, 2626, 2628, 2630, 2632, 2634, 2636, 2638, 2640, 2642, 2644, 2646, 2648, 2650, 2652, 2654, 2656, 2658, 2660, 2662, 2664, 2666, 2668, 2670, 2672, 2674, 2676, 2678, 2680, 2682, 2684, 2686, 2688, 2690, 2692, 2694, 2696, 2698, 2700, 2702, 2704, 2706, 2708, 2710, 2712, 2714, 2716, 2718, 2720, 2722, 2724, 2726, 2728, 2730, 2732, 2734, 2736, 2738, 2740, 2742, 2744, 2746, 2748, 2750, 2752, 2754, 2756, 2758, 2760, 2762, 2764, 2766, 2768, 2770, 2772, 2774, 2776, 2778, 2780, 2782, 2784, 2786, 2788, 2790, 2792, 2794, 2796, 2798, 2800, 2802, 2804, 2806, 2808, 2810, 2812, 2814, 2816, 2818, 2820, 2822, 2824, 2826, 2828, 2830, 2832, 2834, 2836, 2838, 2840, 2842, 2844, 2846, 2848, 2850, 2852, 2854, 2856, 2858, 2860, 2862, 2864, 2866, 2868, 2870, 2872, 2874, 2876, 2878, 2880, 2882, 2884, 2886, 2888, 2890, 2892, 2894, 2896, 2898, 2900, 2902, 2904, 2906, 2908, 2910, 2912, 2914, 2916, 2918, 2920, 2922, 2924, 2926, 2928, 2930, 2932, 2934, 2936, 2938, 2940, 2942, 2944, 2946, 2948, 2950, 2952, 2954, 2956, 2958, 2960, 2962, 2964, 2966, 2968, 2970, 2972, 2974, 2976, 2978, 2980, 2982, 2984, 2986, 2988, 2990, 2992, 2994, 2996, 2998, 3000, 3002, 3004, 3006, 3008, 3010, 3012, 3014, 3016, 3018, 3020, 3022, 3024, 3026, 3028, 3030, 3032, 3034, 3036, 3038, 3040, 3042, 3044, 3046, 3048, 3050, 3052, 3054, 3056, 3058, 3060, 3062, 3064, 3066, 3068, 3070, 3072, 3074, 3076, 3078, 3080, 3082, 3084, 3086, 3088, 3090, 3092, 3094, 3096, 3098, 3100, 3102, 3104, 3106, 3108, 3110, 3112, 3114, 3116, 3118, 3120, 3122, 3124, 3126, 3128, 3130, 3132, 3134, 3136, 3138, 3140, 3142, 3144, 3146, 3148, 3150, 3152, 3154, 3156, 3158, 3160, 3162, 3164, 3166, 3168, 3170, 3172, 3174, 3176, 3178, 3180, 3182, 3184, 3186, 3188, 3190, 3192, 3194, 3196, 3198, 3200, 3202, 3204, 3206, 3208, 3210, 3212, 3214, 3216, 3218, 3220, 3222, 3224, 3226, 3228, 3230, 3232, 3234, 3236, 3238, 3240, 3242, 3244, 3246, 3248, 3250, 3252, 3254, 3256, 3258, 3260, 3262, 3264, 3266, 3268, 3270, 3272, 3274, 3276, 3278, 3280, 3282, 3284, 3286, 3288, 3290, 3292, 3294, 3296, 3298, 3300, 3302, 3304, 3306, 3308, 3310, 3312, 3314, 3316, 3318, 3320, 3322, 3324, 3326, 3328, 3330, 3332, 3334, 3336, 3338, 3340, 3342, 3344, 3346, 3348, 3350, 3352, 3354, 3356, 3358, 3360, 3362, 3364, 3366, 3368, 3370, 3372, 3374, 3376, 3378, 3380, 3382,

3384, 3386, 3388, 3390, 3392, 3394, 3396, 3398, 3400, 3402, 3404, 3406, 3408, 3410, 3412, 3414, 3416, 3418, 3420, 3422, 3424, 3426, 3428, 3430, 3432, 3434, 3436, 3438, 3440, 3442, 3444, 3446, 3448, 3450, 3452, 3454, 3456, 3458, 3460, 3462, 3464, 3466, 3468, 3470, 3472, 3474, 3476, 3478, 3480, 3482, 3484, 3486, 3488, 3490, 3492, 3494, 3496, 3498, 3500, 3502, 3504, 3506, 3508, 3510, 3512, 3514, 3516, 3518, 3520, 3522, 3524, 3526, 3528, 3530, 3532, 3534, 3536, 3538, 3540, 3542, 3544, 3546, 3548, 3550, 3552, 3554, 3556, 3558, 3560, 3562, 3564, 3566, 3568, 3570, 3572, 3574, 3576, 3578, 3580, 3582, 3584, 3586, 3588, 3590, 3592, 3594, 3596, 3598, 3600, 3602, 3604, 3606, 3608, 3610, 3612, 3614, 3616, 3618, 3620, 3622, 3624, 3626, 3628, 3630, 3632, 3634, 3636, 3638, 3640, 3642, 3644, 3646, 3648, 3650, 3652, 3654, 3656, 3658, 3660, 3662, 3664, 3666, 3668, 3670, 3672, 3674, 3676, 3678, 3680, 3682, 3684, 3686, 3688, 3690, 3692, 3694, 3696, 3698, 3700, 3702, 3704, 3706, 3708, 3710, 3712, 3714, 3716, 3718, 3720, 3722, 3724, 3726, 3728, 3730, 3732, 3734, 3736, 3738, 3740, 3742, 3744, 3746, 3748, 3750, 3752, 3754, 3756, 3758, 3760, 3762, 3764, 3766, 3768, 3770, 3772, 3774, 3776, 3778, 3780, 3782, 3784, 3786, 3788, 3790, 3792, 3794, 3796, 3798, 3800, 3802, 3804, 3806, 3808, 3810, 3812, 3814, 3816, 3818, 3820, 3822, 3824, 3826, 3828, 3830, 3832, 3834, 3836, 3838, 3840, 3842, 3844, 3846, 3848, 3850, 3852, 3854, 3856, 3858, 3860, 3862, 3864, 3866, 3868, 3870, 3872, 3874, 3876, 3878, 3880, 3882, 3884, 3886, 3888, 3890, 3892, 3894, 3896, 3898, 3900, 3902, 3904, 3906, 3908, 3910, 3912, 3914, 3916, 3918, 3920, 3922, 3924, 3926, 3928, 3930, 3932, 3934, 3936, 3938, 3940, 3942, 3944, 3946, 3948, 3950, 3952, 3954, 3956, 3958, 3960, 3962, 3964, 3966, 3968, 3970, 3972, 3974, 3976, 3978, 3980, 3982, 3984, 3986, 3988, 3990, 3992, 3994, 3996, 3998, 4000, 4002, 4004, 4006, 4008, 4010, 4012, 4014, 4016, 4018, 4020, 4022, 4024, 4026, 4028, 4030, 4032, 4034, 4036, 4038, 4040, 4042, 4044, 4046, 4048, 4050, 4052, 4054, 4056, 4058, 4060, 4062, 4064, 4066, 4068, 4070, 4072, 4074, 4076, 4078, 4080, 4082, 4084, 4086, 4088, 4090, 4092, 4094, 4096, 4098, 4100, 4102, 4104, 4106, 4108, 4110, 4112, 4114, 4116, 4118, 4120, 4122, 4124, 4126, 4128, 4130, 4132, 4134, 4136, 4138, 4140, 4142, 4144, 4146, 4148, 4150, 4152, 4154, 4156, 4158, 4160, 4162, 4164, 4166, 4168, 4170, 4172, 4174, 4176, 4178, 4180, 4182, 4184, 4186, 4188, 4190, 4192, 4194, 4196, 4198, 4200, 4202, 4204, 4206, 4208, 4210, 4212, 4214, 4216, 4218, 4220, 4222, 4224, 4226, 4228, 4230, 4232, 4234, 4236, 4238, 4240, 4242, 4244, 4246, 4248, 4250, 4252, 4254, 4256, 4258, 4260, 4262, 4264, 4266, 4268, 4270, 4272, 4274, 4276, 4278, 4280, 4282, 4284, 4286, 4288, 4290, 4292, 4294, 4296, 4298, 4300, 4302, 4304, 4306, 4308, 4310, 4312, 4314, 4316, 4318, 4320, 4322, 4324, 4326, 4328, 4330, 4332, 4334, 4336, 4338, 4340, 4342, 4344, 4346, 4348, 4350, 4352, 4354, 4356, 4358, 4360, 4362, 4364, 4366, 4368, 4370, 4372, 4374, 4376, 4378, 4380, 4382, 4384, 4386, 4388, 4390, 4392, 4394, 4396, 4398, 4400, 4402, 4404, 4406, 4408, 4410, 4412, 4414, 4416, 4418, 4420, 4422, 4424, 4426, 4428, 4430, 4432, 4434, 4436, 4438, 4440, 4442, 4444, 4446, 4448, 4450, 4452, 4454, 4456, 4458, 4460, 4462, 4464, 4466, 4468, 4470, 4472, 4474, 4476, 4478, 4480, 4482, 4484, 4486, 4488, 4490, 4492, 4494, 4496, 4498, 4500, 4502, 4504, 4506, 4508, 4510, 4512, 4514, 4516, 4518, 4520, 4522, 4524, 4526, 4528, 4530, 4532, 4534, 4536, 4538, 4540, 4542, 4544, 4546, 4548, 4550, 4552, 4554, 4556, 4558, 4560, 4562, 4564, 4566, 4568, 4570, 4572, 4574, 4576, 4578, 4580, 4582, 4584, 4586, 4588, 4590, 4592, 4594, 4596, 4598, 4600, 4602, 4604, 4606, 4608, 4610, 4612, 4614, 4616, 4618, 4620, 4622, 4624, 4626, 4628, 4630, 4632, 4634, 4636, 4638, 4640, 4642, 4644, 4646, 4648, 4650, 4652, 4654, 4656, 4658, 4660, 4662, 4664, 4666, 4668, 4670, 4672, 4674, 4676, 4678, 4680, 4682, 4684, 4686, 4688, 4690, 4692, 4694, 4696, 4698, 4700, 4702, 4704, 4706, 4708, 4710, 4712, 4714, 4716, 4718, 4720, 4722, 4724, 4726, 4728, 4730, 4732, 4734, 4736, 4738, 4740, 4742, 4744, 4746, 4748, 4750, 4752, 4754, 4756, 4758, 4760, 4762, 4764, 4766, 4768, 4770, 4772, 4774, 4776, 4778, 4780, 4782, 4784, 4786, 4788, 4790, 4792, 4794, 4796, 4798, 4800, 4802, 4804, 4806, 4808, 4810, 4812, 4814, 4816, 4818, 4820, 4822, 4824, 4826, 4828, 4830, 4832, 4834, 4836, 4838, 4840, 4842, 4844, 4846, 4848, 4850, 4852, 4854, 4856, 4858, 4860, 4862, 4864, 4866, 4868, 4870, 4872, 4874, 4876, 4878, 4880, 4882, 4884, 4886, 4888, 4890, 4892, 4894, 4896, 4898, 4900, 4902, 4904, 4906, 4908, 4910, 4912, 4914, 4916, 4918, 4920, 4922, 4924, 4926, 4928, 4930, 4932, 4934, 4936, 4938, 4940, 4942, 4944, 4946, 4948, 4950, 4952, 4954, 4956, 4958, 4960, 4962, 4964, 4966, 4968, 4970, 4972, 4974, 4976, 4978, 4980, 4982, 4984, 4986, 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5002, 5004, 5006, 5008, 5010, 5012, 5014, 5016, 5018, 5020, 5022, 5024, 5026, 5028, 5030, 5032, 5034, 5036, 5038, 5040, 5042, 5044, 5046, 5048, 5050, 5052, 5054, 5056, 5058, 5060, 5062, 5064, 5066, 5068, 5070, 5072, 5074, 5076, 5078, 5080, 5082, 5084, 5086, 5088, 5090, 5092, 5094, 5096, 5098, 5100, 5102, 5104, 5106, 5108, 5110, 5112, 5114, 5116, 5118, 5120, 5122, 5124, 5126, 5128, 5130, 5132, 5134, 5136, 5138, 5140, 5142,

5144, 5146, 5148, 5150, 5152, 5154, 5156, 5158, 5160, 5162, 5164, 5166, 5168, 5170, 5172, 5174, 5176, 5178, 5180, 5182, 5184, 5186, 5188, 5190, 5192, 5194, 5196, 5198, 5200, 5202, 5204, 5206, 5208, 5210, 5212, 5214, 5216, 5218, 5220, 5222, 5224, 5226, 5228, 5230, 5232, 5234, 5236, 5238, 5240, 5242, 5244, 5246, 5248, 5250, 5252, 5254, 5256, 5258, 5260, 5262, 5264, 5266, 5268, 5270, 5272, 5274, 5276, 5278, 5280, 5282, 5284, 5286, 5288, 5290, 5292, 5294, 5296, 5298, 5300, 5302, 5304, 5306, 5308, 5310, 5312, 5314, 5316, 5318, 5320, 5322, 5324, 5326, 5328, 5330, 5332, 5334, 5336, 5338, 5340, 5342, 5344, 5346, 5348, 5350, 5352, 5354, 5356, 5358, 5360, 5362, 5364, 5366, 5368, 5370, 5372, 5374, 5376, 5378, 5380, 5382, 5384, 5386, 5388, 5390, 5392, 5394, 5396, 5398, 5400, 5402, 5404, 5406, 5408, 5410, 5412, 5414, 5416, 5418, 5420, 5422, 5424, 5426, 5428, 5430, 5432, 5434, 5436, 5438, 5440, 5442, 5444, 5446, 5448, 5450, 5452, 5454, 5456, 5458, 5460, 5462, 5464, 5466, 5468, 5470, 5472, 5474, 5476, 5478, 5480, 5482, 5484, 5486, 5488, 5490, 5492, 5494, 5496, 5498, 5500, 5502, 5504, 5506, 5508, 5510, 5512, 5514, 5516, 5518, 5520, 5522, 5524, 5526, 5528, 5530, 5532, 5534, 5536, 5538, 5540, 5542, 5544, 5546, 5548, 5550, 5552, 5554, 5556, 5558, 5560, 5562, 5564, 5566, 5568, 5570, 5572, 5574, 5576, 5578, 5580, 5582, 5584, 5586, 5588, 5590, 5592, 5594, 5596, 5598, 5600, 5602, 5604, 5606, 5608, 5610, 5612, 5614, 5616, 5618, 5620, 5622, 5624, 5626, 5628, 5630, 5632, 5634, 5636, 5638, 5640, 5642.

4. An antibody which binds to a protein according to any one of claims 1 to 3.

5. A nucleic acid molecule which encodes a protein according to any one of claims 1 to 3.

6. A nucleic acid molecule according to claim 5, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, 891, 893, 895, 897, 899, 901, 903, 905, 907, 909, 911, 913, 915, 917, 919, 921, 923, 925, 927, 929, 931, 933, 935, 937, 939, 941, 943, 945, 947, 949, 951, 953, 955, 957, 959, 961, 963, 965, 967, 969, 971, 973, 975, 977, 979, 981, 983, 985, 987, 989, 991, 993, 995, 997, 999, 1001, 1003, 1005, 1007, 1009, 1011, 1013, 1015, 1017, 1019, 1021, 1023, 1025, 1027, 1029, 1031, 1033, 1035, 1037, 1039, 1041, 1043, 1045, 1047, 1049, 1051, 1053, 1055, 1057, 1059, 1061, 1063, 1065, 1067, 1069, 1071, 1073, 1075, 1077, 1079, 1081, 1083, 1085, 1087, 1089, 1091, 1093, 1095, 1097, 1099, 1101, 1103, 1105, 1107, 1109, 1111, 1113, 1115, 1117, 1119, 1121, 1123, 1125, 1127, 1129, 1131, 1133, 1135, 1137, 1139, 1141, 1143, 1145, 1147, 1149, 1151, 1153, 1155, 1157, 1159, 1161, 1163, 1165, 1167, 1169, 1171, 1173, 1175, 1177, 1179, 1181, 1183, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1221, 1223, 1225, 1227, 1229, 1231, 1233, 1235, 1237, 1239, 1241, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, 1273, 1275, 1277, 1279, 1281, 1283, 1285, 1287, 1289, 1291, 1293, 1295, 1297, 1299, 1301, 1303, 1305, 1307, 1309, 1311, 1313, 1315, 1317, 1319, 1321, 1323, 1325, 1327, 1329, 1331, 1333, 1335, 1337, 1339, 1341, 1343, 1345,

1347, 1349, 1351, 1353, 1355, 1357, 1359, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379, 1381, 1383, 1385, 1387, 1389, 1391, 1393, 1395, 1397, 1399, 1401, 1403, 1405, 1407, 1409, 1411, 1413, 1415, 1417, 1419, 1421, 1423, 1425, 1427, 1429, 1431, 1433, 1435, 1437, 1439, 1441, 1443, 1445, 1447, 1449, 1451, 1453, 1455, 1457, 1459, 1461, 1463, 1465, 1467, 1469, 1471, 1473, 1475, 1477, 1479, 1481, 1483, 1485, 1487, 1489, 1491, 1493, 1495, 1497, 1499, 1501, 1503, 1505, 1507, 1509, 1511, 1513, 1515, 1517, 1519, 1521, 1523, 1525, 1527, 1529, 1531, 1533, 1535, 1537, 1539, 1541, 1543, 1545, 1547, 1549, 1551, 1553, 1555, 1557, 1559, 1561, 1563, 1565, 1567, 1569, 1571, 1573, 1575, 1577, 1579, 1581, 1583, 1585, 1587, 1589, 1591, 1593, 1595, 1597, 1599, 1601, 1603, 1605, 1607, 1609, 1611, 1613, 1615, 1617, 1619, 1621, 1623, 1625, 1627, 1629, 1631, 1633, 1635, 1637, 1639, 1641, 1643, 1645, 1647, 1649, 1651, 1653, 1655, 1657, 1659, 1661, 1663, 1665, 1667, 1669, 1671, 1673, 1675, 1677, 1679, 1681, 1683, 1685, 1687, 1689, 1691, 1693, 1695, 1697, 1699, 1701, 1703, 1705, 1707, 1709, 1711, 1713, 1715, 1717, 1719, 1721, 1723, 1725, 1727, 1729, 1731, 1733, 1735, 1737, 1739, 1741, 1743, 1745, 1747, 1749, 1751, 1753, 1755, 1757, 1759, 1761, 1763, 1765, 1767, 1769, 1771, 1773, 1775, 1777, 1779, 1781, 1783, 1785, 1787, 1789, 1791, 1793, 1795, 1797, 1799, 1801, 1803, 1805, 1807, 1809, 1811, 1813, 1815, 1817, 1819, 1821, 1823, 1825, 1827, 1829, 1831, 1833, 1835, 1837, 1839, 1841, 1843, 1845, 1847, 1849, 1851, 1853, 1855, 1857, 1859, 1861, 1863, 1865, 1867, 1869, 1871, 1873, 1875, 1877, 1879, 1881, 1883, 1885, 1887, 1889, 1891, 1893, 1895, 1897, 1899, 1901, 1903, 1905, 1907, 1909, 1911, 1913, 1915, 1917, 1919, 1921, 1923, 1925, 1927, 1929, 1931, 1933, 1935, 1937, 1939, 1941, 1943, 1945, 1947, 1949, 1951, 1953, 1955, 1957, 1959, 1961, 1963, 1965, 1967, 1969, 1971, 1973, 1975, 1977, 1979, 1981, 1983, 1985, 1987, 1989, 1991, 1993, 1995, 1997, 1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019, 2021, 2023, 2025, 2027, 2029, 2031, 2033, 2035, 2037, 2039, 2041, 2043, 2045, 2047, 2049, 2051, 2053, 2055, 2057, 2059, 2061, 2063, 2065, 2067, 2069, 2071, 2073, 2075, 2077, 2079, 2081, 2083, 2085, 2087, 2089, 2091, 2093, 2095, 2097, 2099, 2101, 2103, 2105, 2107, 2109, 2111, 2113, 2115, 2117, 2119, 2121, 2123, 2125, 2127, 2129, 2131, 2133, 2135, 2137, 2139, 2141, 2143, 2145, 2147, 2149, 2151, 2153, 2155, 2157, 2159, 2161, 2163, 2165, 2167, 2169, 2171, 2173, 2175, 2177, 2179, 2181, 2183, 2185, 2187, 2189, 2191, 2193, 2195, 2197, 2199, 2201, 2203, 2205, 2207, 2209, 2211, 2213, 2215, 2217, 2219, 2221, 2223, 2225, 2227, 2229, 2231, 2233, 2235, 2237, 2239, 2241, 2243, 2245, 2247, 2249, 2251, 2253, 2255, 2257, 2259, 2261, 2263, 2265, 2267, 2269, 2271, 2273, 2275, 2277, 2279, 2281, 2283, 2285, 2287, 2289, 2291, 2293, 2295, 2297, 2299, 2301, 2303, 2305, 2307, 2309, 2311, 2313, 2315, 2317, 2319, 2321, 2323, 2325, 2327, 2329, 2331, 2333, 2335, 2337, 2339, 2341, 2343, 2345, 2347, 2349, 2351, 2353, 2355, 2357, 2359, 2361, 2363, 2365, 2367, 2369, 2371, 2373, 2375, 2377, 2379, 2381, 2383, 2385, 2387, 2389, 2391, 2393, 2395, 2397, 2399, 2401, 2403, 2405, 2407, 2409, 2411, 2413, 2415, 2417, 2419, 2421, 2423, 2425, 2427, 2429, 2431, 2433, 2435, 2437, 2439, 2441, 2443, 2445, 2447, 2449, 2451, 2453, 2455, 2457, 2459, 2461, 2463, 2465, 2467, 2469, 2471, 2473, 2475, 2477, 2479, 2481, 2483, 2485, 2487, 2489, 2491, 2493, 2495, 2497, 2499, 2501, 2503, 2505, 2507, 2509, 2511, 2513, 2515, 2517, 2519, 2521, 2523, 2525, 2527, 2529, 2531, 2533, 2535, 2537, 2539, 2541, 2543, 2545, 2547, 2549, 2551, 2553, 2555, 2557, 2559, 2561, 2563, 2565, 2567, 2569, 2571, 2573, 2575, 2577, 2579, 2581, 2583, 2585, 2587, 2589, 2591, 2593, 2595, 2597, 2599, 2601, 2603, 2605, 2607, 2609, 2611, 2613, 2615, 2617, 2619, 2621, 2623, 2625, 2627, 2629, 2631, 2633, 2635, 2637, 2639, 2641, 2643, 2645, 2647, 2649, 2651, 2653, 2655, 2657, 2659, 2661, 2663, 2665, 2667, 2669, 2671, 2673, 2675, 2677, 2679, 2681, 2683, 2685, 2687, 2689, 2691, 2693, 2695, 2697, 2699, 2701, 2703, 2705, 2707, 2709, 2711, 2713, 2715, 2717, 2719, 2721, 2723, 2725, 2727, 2729, 2731, 2733, 2735, 2737, 2739, 2741, 2743, 2745, 2747, 2749, 2751, 2753, 2755, 2757, 2759, 2761, 2763, 2765, 2767, 2769, 2771, 2773, 2775, 2777, 2779, 2781, 2783, 2785, 2787, 2789, 2791, 2793, 2795, 2797, 2799, 2801, 2803, 2805, 2807, 2809, 2811, 2813, 2815, 2817, 2819, 2821, 2823, 2825, 2827, 2829, 2831, 2833, 2835, 2837, 2839, 2841, 2843, 2845, 2847, 2849, 2851, 2853, 2855, 2857, 2859, 2861, 2863, 2865, 2867, 2869, 2871, 2873, 2875, 2877, 2879, 2881, 2883, 2885, 2887, 2889, 2891, 2893, 2895, 2897, 2899, 2901, 2903, 2905, 2907, 2909, 2911, 2913, 2915, 2917, 2919, 2921, 2923, 2925, 2927, 2929, 2931, 2933, 2935, 2937, 2939, 2941, 2943, 2945, 2947, 2949, 2951, 2953, 2955, 2957, 2959, 2961, 2963, 2965, 2967, 2969, 2971, 2973, 2975, 2977, 2979, 2981, 2983, 2985, 2987, 2989, 2991, 2993, 2995, 2997, 2999, 3001, 3003, 3005, 3007, 3009, 3011, 3013, 3015, 3017, 3019, 3021, 3023, 3025, 3027, 3029, 3031, 3033, 3035, 3037, 3039, 3041, 3043, 3045, 3047, 3049, 3051, 3053, 3055, 3057, 3059, 3061, 3063, 3065, 3067, 3069, 3071, 3073, 3075, 3077, 3079, 3081, 3083, 3085, 3087, 3089, 3091, 3093, 3095, 3097, 3099, 3101, 3103, 3105, 3107, 3109, 3111, 3113, 3115, 3117, 3119, 3121, 3123, 3125, 3127, 3129, 3131, 3133, 3135, 3137, 3139, 3141, 3143, 3145, 3147, 3149, 3151, 3153, 3155, 3157, 3159, 3161, 3163, 3165, 3167, 3169, 3171, 3173, 3175, 3177, 3179, 3181, 3183, 3185, 3187, 3189, 3191, 3193,

3195, 3197, 3199, 3201, 3203, 3205, 3207, 3209, 3211, 3213, 3215, 3217, 3219, 3221, 3223, 3225, 3227, 3229, 3231, 3233, 3235, 3237, 3239, 3241, 3243, 3245, 3247, 3249, 3251, 3253, 3255, 3257, 3259, 3261, 3263, 3265, 3267, 3269, 3271, 3273, 3275, 3277, 3279, 3281, 3283, 3285, 3287, 3289, 3291, 3293, 3295, 3297, 3299, 3301, 3303, 3305, 3307, 3309, 3311, 3313, 3315, 3317, 3319, 3321, 3323, 3325, 3327, 3329, 3331, 3333, 3335, 3337, 3339, 3341, 3343, 3345, 3347, 3349, 3351, 3353, 3355, 3357, 3359, 3361, 3363, 3365, 3367, 3369, 3371, 3373, 3375, 3377, 3379, 3381, 3383, 3385, 3387, 3389, 3391, 3393, 3395, 3397, 3399, 3401, 3403, 3405, 3407, 3409, 3411, 3413, 3415, 3417, 3419, 3421, 3423, 3425, 3427, 3429, 3431, 3433, 3435, 3437, 3439, 3441, 3443, 3445, 3447, 3449, 3451, 3453, 3455, 3457, 3459, 3461, 3463, 3465, 3467, 3469, 3471, 3473, 3475, 3477, 3479, 3481, 3483, 3485, 3487, 3489, 3491, 3493, 3495, 3497, 3499, 3501, 3503, 3505, 3507, 3509, 3511, 3513, 3515, 3517, 3519, 3521, 3523, 3525, 3527, 3529, 3531, 3533, 3535, 3537, 3539, 3541, 3543, 3545, 3547, 3549, 3551, 3553, 3555, 3557, 3559, 3561, 3563, 3565, 3567, 3569, 3571, 3573, 3575, 3577, 3579, 3581, 3583, 3585, 3587, 3589, 3591, 3593, 3595, 3597, 3599, 3601, 3603, 3605, 3607, 3609, 3611, 3613, 3615, 3617, 3619, 3621, 3623, 3625, 3627, 3629, 3631, 3633, 3635, 3637, 3639, 3641, 3643, 3645, 3647, 3649, 3651, 3653, 3655, 3657, 3659, 3661, 3663, 3665, 3667, 3669, 3671, 3673, 3675, 3677, 3679, 3681, 3683, 3685, 3687, 3689, 3691, 3693, 3695, 3697, 3699, 3701, 3703, 3705, 3707, 3709, 3711, 3713, 3715, 3717, 3719, 3721, 3723, 3725, 3727, 3729, 3731, 3733, 3735, 3737, 3739, 3741, 3743, 3745, 3747, 3749, 3751, 3753, 3755, 3757, 3759, 3761, 3763, 3765, 3767, 3769, 3771, 3773, 3775, 3777, 3779, 3781, 3783, 3785, 3787, 3789, 3791, 3793, 3795, 3797, 3799, 3801, 3803, 3805, 3807, 3809, 3811, 3813, 3815, 3817, 3819, 3821, 3823, 3825, 3827, 3829, 3831, 3833, 3835, 3837, 3839, 3841, 3843, 3845, 3847, 3849, 3851, 3853, 3855, 3857, 3859, 3861, 3863, 3865, 3867, 3869, 3871, 3873, 3875, 3877, 3879, 3881, 3883, 3885, 3887, 3889, 3891, 3893, 3895, 3897, 3899, 3901, 3903, 3905, 3907, 3909, 3911, 3913, 3915, 3917, 3919, 3921, 3923, 3925, 3927, 3929, 3931, 3933, 3935, 3937, 3939, 3941, 3943, 3945, 3947, 3949, 3951, 3953, 3955, 3957, 3959, 3961, 3963, 3965, 3967, 3969, 3971, 3973, 3975, 3977, 3979, 3981, 3983, 3985, 3987, 3989, 3991, 3993, 3995, 3997, 3999, 4001, 4003, 4005, 4007, 4009, 4011, 4013, 4015, 4017, 4019, 4021, 4023, 4025, 4027, 4029, 4031, 4033, 4035, 4037, 4039, 4041, 4043, 4045, 4047, 4049, 4051, 4053, 4055, 4057, 4059, 4061, 4063, 4065, 4067, 4069, 4071, 4073, 4075, 4077, 4079, 4081, 4083, 4085, 4087, 4089, 4091, 4093, 4095, 4097, 4099, 4101, 4103, 4105, 4107, 4109, 4111, 4113, 4115, 4117, 4119, 4121, 4123, 4125, 4127, 4129, 4131, 4133, 4135, 4137, 4139, 4141, 4143, 4145, 4147, 4149, 4151, 4153, 4155, 4157, 4159, 4161, 4163, 4165, 4167, 4169, 4171, 4173, 4175, 4177, 4179, 4181, 4183, 4185, 4187, 4189, 4191, 4193, 4195, 4197, 4199, 4201, 4203, 4205, 4207, 4209, 4211, 4213, 4215, 4217, 4219, 4221, 4223, 4225, 4227, 4229, 4231, 4233, 4235, 4237, 4239, 4241, 4243, 4245, 4247, 4249, 4251, 4253, 4255, 4257, 4259, 4261, 4263, 4265, 4267, 4269, 4271, 4273, 4275, 4277, 4279, 4281, 4283, 4285, 4287, 4289, 4291, 4293, 4295, 4297, 4299, 4301, 4303, 4305, 4307, 4309, 4311, 4313, 4315, 4317, 4319, 4321, 4323, 4325, 4327, 4329, 4331, 4333, 4335, 4337, 4339, 4341, 4343, 4345, 4347, 4349, 4351, 4353, 4355, 4357, 4359, 4361, 4363, 4365, 4367, 4369, 4371, 4373, 4375, 4377, 4379, 4381, 4383, 4385, 4387, 4389, 4391, 4393, 4395, 4397, 4399, 4401, 4403, 4405, 4407, 4409, 4411, 4413, 4415, 4417, 4419, 4421, 4423, 4425, 4427, 4429, 4431, 4433, 4435, 4437, 4439, 4441, 4443, 4445, 4447, 4449, 4451, 4453, 4455, 4457, 4459, 4461, 4463, 4465, 4467, 4469, 4471, 4473, 4475, 4477, 4479, 4481, 4483, 4485, 4487, 4489, 4491, 4493, 4495, 4497, 4499, 4501, 4503, 4505, 4507, 4509, 4511, 4513, 4515, 4517, 4519, 4521, 4523, 4525, 4527, 4529, 4531, 4533, 4535, 4537, 4539, 4541, 4543, 4545, 4547, 4549, 4551, 4553, 4555, 4557, 4559, 4561, 4563, 4565, 4567, 4569, 4571, 4573, 4575, 4577, 4579, 4581, 4583, 4585, 4587, 4589, 4591, 4593, 4595, 4597, 4599, 4601, 4603, 4605, 4607, 4609, 4611, 4613, 4615, 4617, 4619, 4621, 4623, 4625, 4627, 4629, 4631, 4633, 4635, 4637, 4639, 4641, 4643, 4645, 4647, 4649, 4651, 4653, 4655, 4657, 4659, 4661, 4663, 4665, 4667, 4669, 4671, 4673, 4675, 4677, 4679, 4681, 4683, 4685, 4687, 4689, 4691, 4693, 4695, 4697, 4699, 4701, 4703, 4705, 4707, 4709, 4711, 4713, 4715, 4717, 4719, 4721, 4723, 4725, 4727, 4729, 4731, 4733, 4735, 4737, 4739, 4741, 4743, 4745, 4747, 4749, 4751, 4753, 4755, 4757, 4759, 4761, 4763, 4765, 4767, 4769, 4771, 4773, 4775, 4777, 4779, 4781, 4783, 4785, 4787, 4789, 4791, 4793, 4795, 4797, 4799, 4801, 4803, 4805, 4807, 4809, 4811, 4813, 4815, 4817, 4819, 4821, 4823, 4825, 4827, 4829, 4831, 4833, 4835, 4837, 4839, 4841, 4843, 4845, 4847, 4849, 4851, 4853, 4855, 4857, 4859, 4861, 4863, 4865, 4867, 4869, 4871, 4873, 4875, 4877, 4879, 4881, 4883, 4885, 4887, 4889, 4891, 4893, 4895, 4897, 4899, 4901, 4903, 4905, 4907, 4909, 4911, 4913, 4915, 4917, 4919, 4921, 4923, 4925, 4927, 4929, 4931, 4933, 4935, 4937, 4939, 4941, 4943, 4945, 4947, 4949, 4951, 4953, 4955, 4957, 4959, 4961, 4963, 4965, 4967, 4969, 4971, 4973, 4975, 4977, 4979, 4981, 4983, 4985, 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5001, 5003, 5005, 5007, 5009, 5011, 5013, 5015, 5017, 5019, 5021, 5023, 5025, 5027, 5029, 5031, 5033, 5035, 5037, 5039, 5041,

5043, 5045, 5047, 5049, 5051, 5053, 5055, 5057, 5059, 5061, 5063, 5065, 5067, 5069, 5071, 5073, 5075, 5077, 5079, 5081, 5083, 5085, 5087, 5089, 5091, 5093, 5095, 5097, 5099, 5101, 5103, 5105, 5107, 5109, 5111, 5113, 5115, 5117, 5119, 5121, 5123, 5125, 5127, 5129, 5131, 5133, 5135, 5137, 5139, 5141, 5143, 5145, 5147, 5149, 5151, 5153, 5155, 5157, 5159, 5161, 5163, 5165, 5167, 5169, 5171, 5173, 5175, 5177, 5179, 5181, 5183, 5185, 5187, 5189, 5191, 5193, 5195, 5197, 5199, 5201, 5203, 5205, 5207, 5209, 5211, 5213, 5215, 5217, 5219, 5221, 5223, 5225, 5227, 5229, 5231, 5233, 5235, 5237, 5239, 5241, 5243, 5245, 5247, 5249, 5251, 5253, 5255, 5257, 5259, 5261, 5263, 5265, 5267, 5269, 5271, 5273, 5275, 5277, 5279, 5281, 5283, 5285, 5287, 5289, 5291, 5293, 5295, 5297, 5299, 5301, 5303, 5305, 5307, 5309, 5311, 5313, 5315, 5317, 5319, 5321, 5323, 5325, 5327, 5329, 5331, 5333, 5335, 5337, 5339, 5341, 5343, 5345, 5347, 5349, 5351, 5353, 5355, 5357, 5359, 5361, 5363, 5365, 5367, 5369, 5371, 5373, 5375, 5377, 5379, 5381, 5383, 5385, 5387, 5389, 5391, 5393, 5395, 5397, 5399, 5401, 5403, 5405, 5407, 5409, 5411, 5413, 5415, 5417, 5419, 5421, 5423, 5425, 5427, 5429, 5431, 5433, 5435, 5437, 5439, 5441, 5443, 5445, 5447, 5449, 5451, 5453, 5455, 5457, 5459, 5461, 5463, 5465, 5467, 5469, 5471, 5473, 5475, 5477, 5479, 5481, 5483, 5485, 5487, 5489, 5491, 5493, 5495, 5497, 5499, 5501, 5503, 5505, 5507, 5509, 5511, 5513, 5515, 5517, 5519, 5521, 5523, 5525, 5527, 5529, 5531, 5533, 5535, 5537, 5539, 5541, 5543, 5545, 5547, 5549, 5551, 5553, 5555, 5557, 5559, 5561, 5563, 5565, 5567, 5569, 5571, 5573, 5575, 5577, 5579, 5581, 5583, 5585, 5587, 5589, 5591, 5593, 5595, 5597, 5599, 5601, 5603, 5605, 5607, 5609, 5611, 5613, 5615, 5617, 5619, 5621, 5623, 5625, 5627, 5629, 5631, 5633, 5635, 5637, 5639, 5641.

7. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, 891, 893, 895, 897, 899, 901, 903, 905, 907, 909, 911, 913, 915, 917, 919, 921, 923, 925, 927, 929, 931, 933, 935, 937, 939, 941, 943, 945, 947, 949, 951, 953, 955, 957, 959, 961, 963, 965, 967, 969, 971, 973, 975, 977, 979, 981, 983, 985, 987, 989, 991, 993, 995, 997, 999, 1001, 1003, 1005, 1007, 1009, 1011, 1013, 1015, 1017, 1019, 1021, 1023, 1025, 1027, 1029, 1031, 1033, 1035, 1037, 1039, 1041, 1043, 1045, 1047, 1049, 1051, 1053, 1055, 1057, 1059, 1061, 1063, 1065, 1067, 1069, 1071, 1073, 1075, 1077, 1079, 1081, 1083, 1085, 1087, 1089, 1091, 1093, 1095, 1097, 1099, 1101, 1103, 1105, 1107, 1109, 1111, 1113, 1115, 1117, 1119, 1121, 1123, 1125, 1127, 1129, 1131, 1133, 1135, 1137, 1139, 1141, 1143, 1145, 1147, 1149, 1151, 1153, 1155, 1157, 1159, 1161, 1163, 1165, 1167, 1169, 1171, 1173, 1175, 1177, 1179, 1181, 1183, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1221, 1223, 1225, 1227, 1229, 1231, 1233, 1235, 1237, 1239, 1241, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, 1273, 1275, 1277, 1279, 1281, 1283, 1285, 1287, 1289, 1291, 1293, 1295, 1297, 1299, 1301, 1303, 1305, 1307, 1309, 1311, 1313, 1315, 1317, 1319, 1321, 1323, 1325, 1327, 1329, 1331, 1333, 1335, 1337, 1339,

1341, 1343, 1345, 1347, 1349, 1351, 1353, 1355, 1357, 1359, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379,
1381, 1383, 1385, 1387, 1389, 1391, 1393, 1395, 1397, 1399, 1401, 1403, 1405, 1407, 1409, 1411, 1413, 1415, 1417, 1419,
1421, 1423, 1425, 1427, 1429, 1431, 1433, 1435, 1437, 1439, 1441, 1443, 1445, 1447, 1449, 1451, 1453, 1455, 1457, 1459,
1461, 1463, 1465, 1467, 1469, 1471, 1473, 1475, 1477, 1479, 1481, 1483, 1485, 1487, 1489, 1491, 1493, 1495, 1497, 1499,
1501, 1503, 1505, 1507, 1509, 1511, 1513, 1515, 1517, 1519, 1521, 1523, 1525, 1527, 1529, 1531, 1533, 1535, 1537, 1539,
1541, 1543, 1545, 1547, 1549, 1551, 1553, 1555, 1557, 1559, 1561, 1563, 1565, 1567, 1569, 1571, 1573, 1575, 1577, 1579,
1581, 1583, 1585, 1587, 1589, 1591, 1593, 1595, 1597, 1599, 1601, 1603, 1605, 1607, 1609, 1611, 1613, 1615, 1617, 1619,
1621, 1623, 1625, 1627, 1629, 1631, 1633, 1635, 1637, 1639, 1641, 1643, 1645, 1647, 1649, 1651, 1653, 1655, 1657, 1659,
1661, 1663, 1665, 1667, 1669, 1671, 1673, 1675, 1677, 1679, 1681, 1683, 1685, 1687, 1689, 1691, 1693, 1695, 1697, 1699,
1701, 1703, 1705, 1707, 1709, 1711, 1713, 1715, 1717, 1719, 1721, 1723, 1725, 1727, 1729, 1731, 1733, 1735, 1737, 1739,
1741, 1743, 1745, 1747, 1749, 1751, 1753, 1755, 1757, 1759, 1761, 1763, 1765, 1767, 1769, 1771, 1773, 1775, 1777, 1779,
1781, 1783, 1785, 1787, 1789, 1791, 1793, 1795, 1797, 1799, 1801, 1803, 1805, 1807, 1809, 1811, 1813, 1815, 1817, 1819,
1821, 1823, 1825, 1827, 1829, 1831, 1833, 1835, 1837, 1839, 1841, 1843, 1845, 1847, 1849, 1851, 1853, 1855, 1857, 1859,
1861, 1863, 1865, 1867, 1869, 1871, 1873, 1875, 1877, 1879, 1881, 1883, 1885, 1887, 1889, 1891, 1893, 1895, 1897, 1899,
1901, 1903, 1905, 1907, 1909, 1911, 1913, 1915, 1917, 1919, 1921, 1923, 1925, 1927, 1929, 1931, 1933, 1935, 1937, 1939,
1941, 1943, 1945, 1947, 1949, 1951, 1953, 1955, 1957, 1959, 1961, 1963, 1965, 1967, 1969, 1971, 1973, 1975, 1977, 1979,
1981, 1983, 1985, 1987, 1989, 1991, 1993, 1995, 1997, 1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019,
2021, 2023, 2025, 2027, 2029, 2031, 2033, 2035, 2037, 2039, 2041, 2043, 2045, 2047, 2049, 2051, 2053, 2055, 2057, 2059,
2061, 2063, 2065, 2067, 2069, 2071, 2073, 2075, 2077, 2079, 2081, 2083, 2085, 2087, 2089, 2091, 2093, 2095, 2097, 2099,
2101, 2103, 2105, 2107, 2109, 2111, 2113, 2115, 2117, 2119, 2121, 2123, 2125, 2127, 2129, 2131, 2133, 2135, 2137, 2139,
2141, 2143, 2145, 2147, 2149, 2151, 2153, 2155, 2157, 2159, 2161, 2163, 2165, 2167, 2169, 2171, 2173, 2175, 2177, 2179,
2181, 2183, 2185, 2187, 2189, 2191, 2193, 2195, 2197, 2199, 2201, 2203, 2205, 2207, 2209, 2211, 2213, 2215, 2217, 2219,
2221, 2223, 2225, 2227, 2229, 2231, 2233, 2235, 2237, 2239, 2241, 2243, 2245, 2247, 2249, 2251, 2253, 2255, 2257, 2259,
2261, 2263, 2265, 2267, 2269, 2271, 2273, 2275, 2277, 2279, 2281, 2283, 2285, 2287, 2289, 2291, 2293, 2295, 2297, 2299,
2301, 2303, 2305, 2307, 2309, 2311, 2313, 2315, 2317, 2319, 2321, 2323, 2325, 2327, 2329, 2331, 2333, 2335, 2337, 2339,
2341, 2343, 2345, 2347, 2349, 2351, 2353, 2355, 2357, 2359, 2361, 2363, 2365, 2367, 2369, 2371, 2373, 2375, 2377, 2379,
2381, 2383, 2385, 2387, 2389, 2391, 2393, 2395, 2397, 2399, 2401, 2403, 2405, 2407, 2409, 2411, 2413, 2415, 2417, 2419,
2421, 2423, 2425, 2427, 2429, 2431, 2433, 2435, 2437, 2439, 2441, 2443, 2445, 2447, 2449, 2451, 2453, 2455, 2457, 2459,
2461, 2463, 2465, 2467, 2469, 2471, 2473, 2475, 2477, 2479, 2481, 2483, 2485, 2487, 2489, 2491, 2493, 2495, 2497, 2499,
2501, 2503, 2505, 2507, 2509, 2511, 2513, 2515, 2517, 2519, 2521, 2523, 2525, 2527, 2529, 2531, 2533, 2535, 2537, 2539,
2541, 2543, 2545, 2547, 2549, 2551, 2553, 2555, 2557, 2559, 2561, 2563, 2565, 2567, 2569, 2571, 2573, 2575, 2577, 2579,
2581, 2583, 2585, 2587, 2589, 2591, 2593, 2595, 2597, 2599, 2601, 2603, 2605, 2607, 2609, 2611, 2613, 2615, 2617, 2619,
2621, 2623, 2625, 2627, 2629, 2631, 2633, 2635, 2637, 2639, 2641, 2643, 2645, 2647, 2649, 2651, 2653, 2655, 2657, 2659,
2661, 2663, 2665, 2667, 2669, 2671, 2673, 2675, 2677, 2679, 2681, 2683, 2685, 2687, 2689, 2691, 2693, 2695, 2697, 2699,
2701, 2703, 2705, 2707, 2709, 2711, 2713, 2715, 2717, 2719, 2721, 2723, 2725, 2727, 2729, 2731, 2733, 2735, 2737, 2739,
2741, 2743, 2745, 2747, 2749, 2751, 2753, 2755, 2757, 2759, 2761, 2763, 2765, 2767, 2769, 2771, 2773, 2775, 2777, 2779,
2781, 2783, 2785, 2787, 2789, 2791, 2793, 2795, 2797, 2799, 2801, 2803, 2805, 2807, 2809, 2811, 2813, 2815, 2817, 2819,
2821, 2823, 2825, 2827, 2829, 2831, 2833, 2835, 2837, 2839, 2841, 2843, 2845, 2847, 2849, 2851, 2853, 2855, 2857, 2859,
2861, 2863, 2865, 2867, 2869, 2871, 2873, 2875, 2877, 2879, 2881, 2883, 2885, 2887, 2889, 2891, 2893, 2895, 2897, 2899,
2901, 2903, 2905, 2907, 2909, 2911, 2913, 2915, 2917, 2919, 2921, 2923, 2925, 2927, 2929, 2931, 2933, 2935, 2937, 2939,
2941, 2943, 2945, 2947, 2949, 2951, 2953, 2955, 2957, 2959, 2961, 2963, 2965, 2967, 2969, 2971, 2973, 2975, 2977, 2979,
2981, 2983, 2985, 2987, 2989, 2991, 2993, 2995, 2997, 2999, 3001, 3003, 3005, 3007, 3009, 3011, 3013, 3015, 3017, 3019,
3021, 3023, 3025, 3027, 3029, 3031, 3033, 3035, 3037, 3039, 3041, 3043, 3045, 3047, 3049, 3051, 3053, 3055, 3057, 3059,
3061, 3063, 3065, 3067, 3069, 3071, 3073, 3075, 3077, 3079, 3081, 3083, 3085, 3087, 3089, 3091, 3093, 3095, 3097, 3099,

3101, 3103, 3105, 3107, 3109, 3111, 3113, 3115, 3117, 3119, 3121, 3123, 3125, 3127, 3129, 3131, 3133, 3135, 3137, 3139, 3141, 3143, 3145, 3147, 3149, 3151, 3153, 3155, 3157, 3159, 3161, 3163, 3165, 3167, 3169, 3171, 3173, 3175, 3177, 3179, 3181, 3183, 3185, 3187, 3189, 3191, 3193, 3195, 3197, 3199, 3201, 3203, 3205, 3207, 3209, 3211, 3213, 3215, 3217, 3219, 3221, 3223, 3225, 3227, 3229, 3231, 3233, 3235, 3237, 3239, 3241, 3243, 3245, 3247, 3249, 3251, 3253, 3255, 3257, 3259, 3261, 3263, 3265, 3267, 3269, 3271, 3273, 3275, 3277, 3279, 3281, 3283, 3285, 3287, 3289, 3291, 3293, 3295, 3297, 3299, 3301, 3303, 3305, 3307, 3309, 3311, 3313, 3315, 3317, 3319, 3321, 3323, 3325, 3327, 3329, 3331, 3333, 3335, 3337, 3339, 3341, 3343, 3345, 3347, 3349, 3351, 3353, 3355, 3357, 3359, 3361, 3363, 3365, 3367, 3369, 3371, 3373, 3375, 3377, 3379, 3381, 3383, 3385, 3387, 3389, 3391, 3393, 3395, 3397, 3399, 3401, 3403, 3405, 3407, 3409, 3411, 3413, 3415, 3417, 3419, 3421, 3423, 3425, 3427, 3429, 3431, 3433, 3435, 3437, 3439, 3441, 3443, 3445, 3447, 3449, 3451, 3453, 3455, 3457, 3459, 3461, 3463, 3465, 3467, 3469, 3471, 3473, 3475, 3477, 3479, 3481, 3483, 3485, 3487, 3489, 3491, 3493, 3495, 3497, 3499, 3501, 3503, 3505, 3507, 3509, 3511, 3513, 3515, 3517, 3519, 3521, 3523, 3525, 3527, 3529, 3531, 3533, 3535, 3537, 3539, 3541, 3543, 3545, 3547, 3549, 3551, 3553, 3555, 3557, 3559, 3561, 3563, 3565, 3567, 3569, 3571, 3573, 3575, 3577, 3579, 3581, 3583, 3585, 3587, 3589, 3591, 3593, 3595, 3597, 3599, 3601, 3603, 3605, 3607, 3609, 3611, 3613, 3615, 3617, 3619, 3621, 3623, 3625, 3627, 3629, 3631, 3633, 3635, 3637, 3639, 3641, 3643, 3645, 3647, 3649, 3651, 3653, 3655, 3657, 3659, 3661, 3663, 3665, 3667, 3669, 3671, 3673, 3675, 3677, 3679, 3681, 3683, 3685, 3687, 3689, 3691, 3693, 3695, 3697, 3699, 3701, 3703, 3705, 3707, 3709, 3711, 3713, 3715, 3717, 3719, 3721, 3723, 3725, 3727, 3729, 3731, 3733, 3735, 3737, 3739, 3741, 3743, 3745, 3747, 3749, 3751, 3753, 3755, 3757, 3759, 3761, 3763, 3765, 3767, 3769, 3771, 3773, 3775, 3777, 3779, 3781, 3783, 3785, 3787, 3789, 3791, 3793, 3795, 3797, 3799, 3801, 3803, 3805, 3807, 3809, 3811, 3813, 3815, 3817, 3819, 3821, 3823, 3825, 3827, 3829, 3831, 3833, 3835, 3837, 3839, 3841, 3843, 3845, 3847, 3849, 3851, 3853, 3855, 3857, 3859, 3861, 3863, 3865, 3867, 3869, 3871, 3873, 3875, 3877, 3879, 3881, 3883, 3885, 3887, 3889, 3891, 3893, 3895, 3897, 3899, 3901, 3903, 3905, 3907, 3909, 3911, 3913, 3915, 3917, 3919, 3921, 3923, 3925, 3927, 3929, 3931, 3933, 3935, 3937, 3939, 3941, 3943, 3945, 3947, 3949, 3951, 3953, 3955, 3957, 3959, 3961, 3963, 3965, 3967, 3969, 3971, 3973, 3975, 3977, 3979, 3981, 3983, 3985, 3987, 3989, 3991, 3993, 3995, 3997, 3999, 4001, 4003, 4005, 4007, 4009, 4011, 4013, 4015, 4017, 4019, 4021, 4023, 4025, 4027, 4029, 4031, 4033, 4035, 4037, 4039, 4041, 4043, 4045, 4047, 4049, 4051, 4053, 4055, 4057, 4059, 4061, 4063, 4065, 4067, 4069, 4071, 4073, 4075, 4077, 4079, 4081, 4083, 4085, 4087, 4089, 4091, 4093, 4095, 4097, 4099, 4101, 4103, 4105, 4107, 4109, 4111, 4113, 4115, 4117, 4119, 4121, 4123, 4125, 4127, 4129, 4131, 4133, 4135, 4137, 4139, 4141, 4143, 4145, 4147, 4149, 4151, 4153, 4155, 4157, 4159, 4161, 4163, 4165, 4167, 4169, 4171, 4173, 4175, 4177, 4179, 4181, 4183, 4185, 4187, 4189, 4191, 4193, 4195, 4197, 4199, 4201, 4203, 4205, 4207, 4209, 4211, 4213, 4215, 4217, 4219, 4221, 4223, 4225, 4227, 4229, 4231, 4233, 4235, 4237, 4239, 4241, 4243, 4245, 4247, 4249, 4251, 4253, 4255, 4257, 4259, 4261, 4263, 4265, 4267, 4269, 4271, 4273, 4275, 4277, 4279, 4281, 4283, 4285, 4287, 4289, 4291, 4293, 4295, 4297, 4299, 4301, 4303, 4305, 4307, 4309, 4311, 4313, 4315, 4317, 4319, 4321, 4323, 4325, 4327, 4329, 4331, 4333, 4335, 4337, 4339, 4341, 4343, 4345, 4347, 4349, 4351, 4353, 4355, 4357, 4359, 4361, 4363, 4365, 4367, 4369, 4371, 4373, 4375, 4377, 4379, 4381, 4383, 4385, 4387, 4389, 4391, 4393, 4395, 4397, 4399, 4401, 4403, 4405, 4407, 4409, 4411, 4413, 4415, 4417, 4419, 4421, 4423, 4425, 4427, 4429, 4431, 4433, 4435, 4437, 4439, 4441, 4443, 4445, 4447, 4449, 4451, 4453, 4455, 4457, 4459, 4461, 4463, 4465, 4467, 4469, 4471, 4473, 4475, 4477, 4479, 4481, 4483, 4485, 4487, 4489, 4491, 4493, 4495, 4497, 4499, 4501, 4503, 4505, 4507, 4509, 4511, 4513, 4515, 4517, 4519, 4521, 4523, 4525, 4527, 4529, 4531, 4533, 4535, 4537, 4539, 4541, 4543, 4545, 4547, 4549, 4551, 4553, 4555, 4557, 4559, 4561, 4563, 4565, 4567, 4569, 4571, 4573, 4575, 4577, 4579, 4581, 4583, 4585, 4587, 4589, 4591, 4593, 4595, 4597, 4599, 4601, 4603, 4605, 4607, 4609, 4611, 4613, 4615, 4617, 4619, 4621, 4623, 4625, 4627, 4629, 4631, 4633, 4635, 4637, 4639, 4641, 4643, 4645, 4647, 4649, 4651, 4653, 4655, 4657, 4659, 4661, 4663, 4665, 4667, 4669, 4671, 4673, 4675, 4677, 4679, 4681, 4683, 4685, 4687, 4689, 4691, 4693, 4695, 4697, 4699, 4701, 4703, 4705, 4707, 4709, 4711, 4713, 4715, 4717, 4719, 4721, 4723, 4725, 4727, 4729, 4731, 4733, 4735, 4737, 4739, 4741, 4743, 4745, 4747, 4749, 4751, 4753, 4755, 4757, 4759, 4761, 4763, 4765, 4767, 4769, 4771, 4773, 4775, 4777, 4779, 4781, 4783, 4785, 4787, 4789, 4791, 4793, 4795, 4797, 4799, 4801, 4803, 4805, 4807, 4809, 4811, 4813, 4815, 4817, 4819, 4821, 4823, 4825, 4827, 4829, 4831, 4833, 4835, 4837, 4839, 4841, 4843, 4845, 4847, 4849, 4851, 4853, 4855, 4857, 4859,

4861, 4863, 4865, 4867, 4869, 4871, 4873, 4875, 4877, 4879, 4881, 4883, 4885, 4887, 4889, 4891, 4893, 4895, 4897, 4899, 4901, 4903, 4905, 4907, 4909, 4911, 4913, 4915, 4917, 4919, 4921, 4923, 4925, 4927, 4929, 4931, 4933, 4935, 4937, 4939, 4941, 4943, 4945, 4947, 4949, 4951, 4953, 4955, 4957, 4959, 4961, 4963, 4965, 4967, 4969, 4971, 4973, 4975, 4977, 4979, 4981, 4983, 4985, 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5001, 5003, 5005, 5007, 5009, 5011, 5013, 5015, 5017, 5019, 5021, 5023, 5025, 5027, 5029, 5031, 5033, 5035, 5037, 5039, 5041, 5043, 5045, 5047, 5049, 5051, 5053, 5055, 5057, 5059, 5061, 5063, 5065, 5067, 5069, 5071, 5073, 5075, 5077, 5079, 5081, 5083, 5085, 5087, 5089, 5091, 5093, 5095, 5097, 5099, 5101, 5103, 5105, 5107, 5109, 5111, 5113, 5115, 5117, 5119, 5121, 5123, 5125, 5127, 5129, 5131, 5133, 5135, 5137, 5139, 5141, 5143, 5145, 5147, 5149, 5151, 5153, 5155, 5157, 5159, 5161, 5163, 5165, 5167, 5169, 5171, 5173, 5175, 5177, 5179, 5181, 5183, 5185, 5187, 5189, 5191, 5193, 5195, 5197, 5199, 5201, 5203, 5205, 5207, 5209, 5211, 5213, 5215, 5217, 5219, 5221, 5223, 5225, 5227, 5229, 5231, 5233, 5235, 5237, 5239, 5241, 5243, 5245, 5247, 5249, 5251, 5253, 5255, 5257, 5259, 5261, 5263, 5265, 5267, 5269, 5271, 5273, 5275, 5277, 5279, 5281, 5283, 5285, 5287, 5289, 5291, 5293, 5295, 5297, 5299, 5301, 5303, 5305, 5307, 5309, 5311, 5313, 5315, 5317, 5319, 5321, 5323, 5325, 5327, 5329, 5331, 5333, 5335, 5337, 5339, 5341, 5343, 5345, 5347, 5349, 5351, 5353, 5355, 5357, 5359, 5361, 5363, 5365, 5367, 5369, 5371, 5373, 5375, 5377, 5379, 5381, 5383, 5385, 5387, 5389, 5391, 5393, 5395, 5397, 5399, 5401, 5403, 5405, 5407, 5409, 5411, 5413, 5415, 5417, 5419, 5421, 5423, 5425, 5427, 5429, 5431, 5433, 5435, 5437, 5439, 5441, 5443, 5445, 5447, 5449, 5451, 5453, 5455, 5457, 5459, 5461, 5463, 5465, 5467, 5469, 5471, 5473, 5475, 5477, 5479, 5481, 5483, 5485, 5487, 5489, 5491, 5493, 5495, 5497, 5499, 5501, 5503, 5505, 5507, 5509, 5511, 5513, 5515, 5517, 5519, 5521, 5523, 5525, 5527, 5529, 5531, 5533, 5535, 5537, 5539, 5541, 5543, 5545, 5547, 5549, 5551, 5553, 5555, 5557, 5559, 5561, 5563, 5565, 5567, 5569, 5571, 5573, 5575, 5577, 5579, 5581, 5583, 5585, 5587, 5589, 5591, 5593, 5595, 5597, 5599, 5601, 5603, 5605, 5607, 5609, 5611, 5613, 5615, 5617, 5619, 5621, 5623, 5625, 5627, 5629, 5631, 5633, 5635, 5637, 5639, 5641.

8. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 5 to 7.

9. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 5 to 8.

10. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 5 to 9 under high stringency conditions.

11. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.

12. A composition according to claim 11 being a vaccine composition or a diagnostic composition.

13. A composition according to claim 11 or claim 12 for use as a pharmaceutical.

14. The use of a composition according to claim 13 in the manufacture of a medicament for the treatment or prevention of infection due to staphylococcus bacteria, particularly *S.aureus*.

15. The use of claim 14, wherein the disease is sepsis.

16. A method of treating a patient, comprising administering to the patient a therapeutically effective amount of a composition according to claim 13.

17. A kit comprising primers for amplifying a target sequence contained within a Staphylococcus nucleic acid sequence, the kit comprising a first primer and a second primer, wherein the first primer is substantially complementary to said target sequence and the second primer is substantially complementary to a complement of said target sequence, wherein the parts

of said primers which have substantial complementarity define the termini of the target sequence to be amplified, and wherein the first and/or second primer is a nucleic acid according to any one of claims 5 to 10.

18. A kit comprising first and second single-stranded oligonucleotides which allow amplification of a Staphylococcus template nucleic acid sequence contained in a single- or double-stranded nucleic acid (or mixture thereof), wherein: (a) the first oligonucleotide comprises a primer sequence which is substantially complementary to said template nucleic acid sequence; (b) the second oligonucleotide comprises a primer sequence which is substantially complementary to the complement of said template nucleic acid sequence; (c) the first oligonucleotide and/or the second oligonucleotide comprise(s) sequence which is not complementary to said template nucleic acid; (d) said primer sequences define the termini of the template sequence to be amplified; and (e) the first and/or second oligonucleotide is a nucleic acid according to any one of claims 5 to 10.

19. A hybrid protein represented by the formula NH_2-A-[-X-L-]$_n$-B-COOH, wherein X is the amino acid sequence of a protein according to claim 1, claim 2 or claim 3, L is an optional linker amino acid sequence, A is an optional N-terminal amino acid sequence, B is an optional C-terminal amino acid sequence, and n is an integer greater than 1.

20. An assay comprising the steps of contacting a test compound with a protein according to any one of claims 1 to 3, and determining whether the test compound binds to said protein.

21. The composition of claim 13 further comprising one or more of the following antigens:
- a protein antigen from *Helicobacter pylori* ;
- a protein antigen from *N.meningitidis* serogroup B;
- an outer-membrane vesicle (OMV) preparation from *N.meningitidis*;
- a saccharide antigen from *N.meningitidis* serogroup A, C, W135 and/or Y;
- a saccharide antigen from *Streptococcus pneumoniae*;
- an antigen from hepatitis A virus;
- an antigen from hepatitis B virus;
- an antigen from hepatitis C virus;
- an antigen from *Bordetella pertussis*;
- a diphtheria antigen;
- a tetanus antigen;
- a saccharide antigen from *Haemophilus influenzae* B.
- an antigen from *N.gonorrhoeae;*
- an antigen from *Chlamydia pneumoniae;*
- an antigen from *Streptococcus agalactiae*;
- an antigen from *Streptococcus pyogenes* ;
- an antigen from *Chlamydia trachomatis* ;
- an antigen from *Porphyromonas gingivalis*;
- polio antigen(s);
- rabies antigen(s);
- measles, mumps and/or rubella antigens;
- influenza antigen(s);
- an antigen from Moraxella catarrhalis; and/or

22. A composition comprising two or more proteins of any one of claims 1 to 3.