



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2375998 A1 2000/12/21

(21) **2 375 998**

(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 1999/06/11
(87) Date publication PCT/PCT Publication Date: 2000/12/21
(85) Entrée phase nationale/National Entry: 2001/11/30
(86) N° demande PCT/PCT Application No.: US 1999/013024
(87) N° publication PCT/PCT Publication No.: 2000/077196

(51) Cl.Int.⁶/Int.Cl.⁶ C12N 15/11, C12N 15/85, C12N 15/63,
C07K 16/00, C07K 14/00

(71) Demandeur/Applicant:
THE TRUSTEES OF TUFTS COLLEGE, US

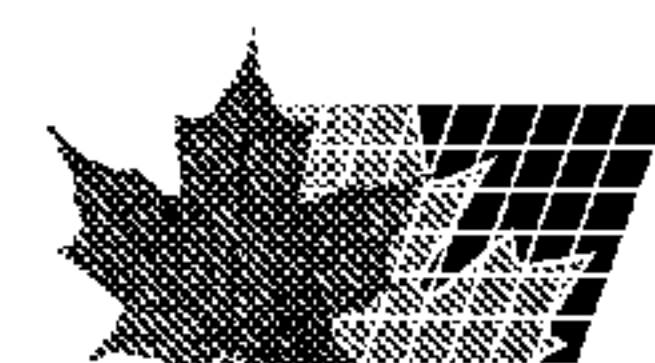
(72) Inventeur/Inventor:
GOLDBERG, EDWARD B., US

(74) Agent: OSLER, HOSKIN & HARCOURT LLP

(54) Titre : GENE ET SEQUENCES DE PROTEINES DU GENE 35 DE PHAGE T4
(54) Title: GENE AND PROTEIN SEQUENCES OF PHAGE T4 GENE 35

(57) **Abrégé/Abstract:**

The present invention relates to nucleotide sequences of gp35 genes and amino acid sequences of their encoded proteins, as well as derivatives and analogs thereof, and antibodies thereto. The present invention further relates to the use of nucleotide sequences of bacteriophage T4 gene 35 and amino acid sequences of its encoded protein, as well as derivatives, variants, and analogs thereof in the construction of nanostructures.



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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 December 2000 (21.12.2000)

PCT

(10) International Publication Number
WO 00/77196 A1

(51) International Patent Classification⁶: **C12N 15/11**,
15/63, 15/85, C07K 14/00, 16/00

(21) International Application Number: PCT/US99/13024

(22) International Filing Date: 11 June 1999 (11.06.1999)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (*for all designated States except US*): **THE TRUSTEES OF TUFTS COLLEGE** [US/US]; Packard Hall, Medford, MA 02155 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **GOLDBERG, Edward, B.** [US/US]; 494 Ward Street, Newton, MA 02459 (US).

(74) Agents: **ANTLER, Adriane, M.** et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, 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, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, 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), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *With international search 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.

(54) Title: GENE AND PROTEIN SEQUENCES OF PHAGE T4 GENE 35

(57) Abstract: The present invention relates to nucleotide sequences of gp35 genes and amino acid sequences of their encoded proteins, as well as derivatives and analogs thereof, and antibodies thereto. The present invention further relates to the use of nucleotide sequences of bacteriophage T4 gene 35 and amino acid sequences of its encoded protein, as well as derivatives, variants, and analogs thereof in the construction of nanostructures.

WO 00/77196 A1

GENE AND PROTEIN SEQUENCES OF PHAGE T4 *gene 35*

1. INTRODUCTION

The present invention relates to nucleotide sequences of bacteriophage T4 *gene 35* and amino acid sequences of its encoded protein, as well as derivatives and analogs thereof and antibodies thereto. The present invention further relates to the use of nucleic acids encoding bacteriophage T4 *gene 35* and its encoded protein, as well as derivatives, and analogs thereof, in the construction of nanostructures, *i.e.*, nanometer sized structures useful in the construction of microscopic and macroscopic structures.

2. BACKGROUND OF THE INVENTION

Bacteriophage, viruses that attack bacteria, are generally composed of a protein coat which surrounds genetic material. Bacteriophage T4, a T-even phage, consists of an icosahedron shaped head which contains DNA, a tail (a hollow cylinder of contractile protein) which serves as an injection tube of the DNA and tail fiber appendages which emanate from the base of the tail. The tail fibers serve to attach the phage to the bacterial surface in a process known as adsorption.

Bacteriophage T4 tail fiber is composed of four non-covalently joined parts in a stiff, heat stable, protease resistant structure. This structure can be represented schematically as follows (N= amino terminus, C= carboxy terminus): N[gp34 homooligomer]C - N[gp35]C - N[gp36 homooligomer]C - N[gp37 homooligomer]C. The gp34 homooligomer ("P34"), gp36 homooligomer ("P36"), and gp37 homooligomer ("P37") are rod-shaped structures in which two identical β sheets, oriented in the same direction, are fused face-to-face by hydrophobic interactions between the sheets juxtaposed with a 180° rotational axis of symmetry through the long axis of the rod. gp35, by contrast, is a monomeric polypeptide that attaches specifically first to the N-terminal region of the P36 homooligomer and then to the C-terminus of the P34 homooligomer and forms a joint between these two rods having an average angle of $137^\circ (\pm 7^\circ)$ or $156^\circ (\pm 12^\circ)$.

The self assembly of the tail fiber is regulated by a predetermined order based on the interaction of specific protein subunits whereby structural maturation caused by formation of the first subassembly permits interaction with new (previously disallowed) subunits. During T4 infection of *E. coli*, gp37 (the monomeric 109 Kda translation product of *gene 37*) forms the homooligomer P37, with the aid of 2 accessory (chaperon) proteins, gp57 and

gp38; this process is believed to initiate near the C-terminus of gp37. Once P37 is formed, the N-terminus of P37 initiates the oligomerization of two gp36 molecules of 23 Kda each, in a butt-end joint to form the P36 homooligomer rod. The N-terminus of P36 then attaches to the carboxy terminal region of a gp35 monomer; this interaction stabilizes P36 and forms
5 the flexible angle joint of the tail fiber. The amino terminal region of gp35 then attaches to the C-terminus of P34 (the homooligomerization of which requires the chaperon protein gp57). This regulation of self assembly of the tail fiber by a predetermined, ordered interaction of specific subunits results in the production of a structure of exact specifications from a random mixture of the tail fiber subunit components.

10 While the strength of most metallic and ceramic based materials derives from the theoretical bonding strengths between their component molecules and crystallite surfaces, it is significantly limited by flaws in their crystal or glass-like structures. These flaws are usually inherent in the raw materials themselves or developed during fabrication and are often expanded due to exposure to environmental stresses.

15 The emerging field of nanotechnology has made the limitations of traditional materials more critical. The ability to design and produce very small structures (*i.e.*, of nanometer dimensions) that can serve complex functions depends upon the use of appropriate materials that can be manipulated in predictable and reproducible ways, and that have the properties required for each novel application.

20 Biological systems serve as a paradigm for sophisticated nanostructures. Living cells fabricate proteins and combine them into structures, such as bacteriophage tail fibers, that are perfectly formed and can resist damage in their normal environment. In some cases, such as with bacteriophage tail fibers, these structures are created by a process of self-assembly, the instructions for which are built into the component polypeptides. These
25 natural proteins are also subject to proofreading processes that insure a high degree of quality control. Advantages of using natural proteins to construct nanostructures are that the resulting structures are stiff, strong, stable in aqueous media, heat resistant, protease resistant, and can be rendered biodegradable. Additionally, large quantities of nanostructure parts and subassemblies can be easily fabricated in microorganisms and stored and used as
30 needed.

There is a need in the art for methods and compositions that exploit these unique features of proteins to form constituents of synthetic nanostructures. The need is to design

materials that have properties which can be tailored to suit the particular requirements of nanometer-scale technology. Moreover, since the subunits of most macrostructural materials, ceramics, metals, fibers, etc., are based on the bonding of nanostructural subunits, the fabrication of appropriate subunits without flaws and of exact dimensions and
5 uniformity should improve the strength and consistency of these macrostructures because the surfaces are more regular and can interact more closely over an extended area than larger, more heterogeneous material.

The use of bacteriophage tail fiber components in the construction of nanostructures is further described in PCT Publication WO 96/11947, dated April 25, 1996, the contents of
10 which are incorporated herein in its entirety.

Phage T4 *gp35* is located between genes *gp34* and *gp36*. A sequence for *gp35* is available on the NCBI database (NCBI.NIH.GOV) within the sequence T4g34-t (bases 4188-5075). The T4g34-t sequence reveals that *gene 35* has an open reading frame, ORF35, that is predicted to encode a protein having a molecular weight of 32,334 Daltons.
15 The NCBI database also predicts an open reading frame, ORF34.1, that extends 241 nucleotides between genes *gp34* and *gp35*, and encodes a deduced protein having a molecular weight of 7,334 Daltons (in a different reading frame from ORF35).

The discrepancy between the *gp35* molecular weight of 32,334 Daltons predicted by the NCBI sequence and that of 39,000-40,000 Daltons reported from SDS-polyacrylamide
20 gel electrophoresis (SDS-PAGE) analysis, has previously been acknowledged (Karam, J. (ed.), 1994, Molecular Biology of Bacteriophage T4, ASM Press, Wash. D.C., pp. 491-514 at 514).

Citation of a reference hereinabove shall not be construed as an admission that such reference is prior art to the present invention.

25

3. SUMMARY OF THE INVENTION

The present invention relates to nucleotide sequences of bacteriophage T4 *gene 35*, and amino acid sequences of the encoded bacteriophage T4 *gene 35* protein, as well as derivatives (*e.g.*, fragments) and analogs thereof, and antibodies thereto. The present
30 invention further relates to nucleic acids hybridizable to or complementary to the foregoing nucleotide sequences, as well as equivalent nucleic acid sequences encoding a bacteriophage T4 *gene 35* protein.

The present invention also relates to expression vectors encoding a bacteriophage T4 gene 35 protein, derivatives or analogs thereof, as well as host cells containing the expression vectors encoding the bacteriophage T4 gene 35 protein, derivative or analog thereof. As used herein, "*gene 35 (gp35)*" shall be used with reference to the bacteriophage
5 T4 *gene 35*, whereas "*gene 35 (gp35)*" shall be used with reference to the protein product of bacteriophage T4 *gene 35*.

The present invention also relates to methods of production of the gp35 proteins, derivatives and analogs, such as, for example, by recombinant means.

The invention further relates to gp35 proteins, derivatives (*e.g.*, fragments), and
10 analogs having an angle joint domain that has been modified so as to form average angles different from the natural average angle of $137^{\circ} (\pm 7^{\circ})$ or $156^{\circ} (\pm 12^{\circ})$.

The invention also relates to gp35 proteins, derivatives and analogs which exhibit thermolabile interactions with tail fiber binding partners.

The invention further relates to gp35 derivatives and analogs which are functionally
15 active, *i.e.*, they are capable of displaying one or more known functional activities associated with a full-length (wild-type) gp35 protein. Such functional activities include, but are not limited to, antigenicity [ability to bind (or compete with gp35 for binding) to an anti-gp35 antibody], immunogenicity (ability to generate antibody which binds to gp35), and ability to bind (or compete with gp35 for binding) to a ligand for gp35, and ability to
20 multimerize with other phage products such as P34 and/or P36.

The gp35 protein, derivative or analogs thereof disclosed herein may be used for the production of anti-gp35 antibodies which antibodies may be used diagnostically in immunoassays for the detection or measurement of gp35 protein.

The invention also relates to fragments (and derivatives and analogs thereof) of gp35
25 which comprise one or more domains of a gp35 protein, *e.g.*, the P34 or P36 binding domain, and/or retain the antigenicity of a gp35 protein (*i.e.*, are able to be bound by an anti-gp35 antibody).

The present invention further relates to the use of nucleotide sequences of *gp35* and its encoded amino acid sequence in the construction of nanostructures, *i.e.*, nanometer sized
30 structures useful in the construction of microscopic and macroscopic structures.

4. DESCRIPTION OF THE FIGURES

Figures 1A-1B. T4 bacteriophage. Schematic representation of the T4 bacteriophage particle (Figure 1A), and a schematic representation of the bacteriophage T4 tail fiber (Figure 1B).

Figure 2. Sequence of bacteriophage T4 gp35. The nucleotide (SEQ ID NO:1) and deduced amino acid (SEQ ID NO:2) sequences of bacteriophage T4 gp35. The gp35 protein sequence shown in Figure 3 (encoded by nucleotides 4,127-5,011 of Figure 3) lacks amino acid numbers 1-77 of Figure 2. Amino acid numbers 1-7, 18-56 and 65 of Figure 2 appear as part of the ORF34.1 sequence in Figure 3 (encoded by nucleotides 3,894-4,088 of Figure 3).

Figure 3. NCBI database sequence containing bacteriophage T4 *gene 34*, *gene 35* (with errors), *gene 36* and *gene 37*. The nucleotide sequence containing *gene 34*, *gene 35* and *gene 36* (SEQ ID NO:3) and the amino acids encoding the gene products of *gene 34* (SEQ ID NO:4; ORF 34.1, SEQ ID NO:5) *gene 35* (SEQ ID NO:6), *GENE 36* (SEQ ID NO:7) and *gene 37* (SEQ ID NO:8).

5. DETAILED DESCRIPTION OF THE INVENTION

The present inventor has discovered that significant errors are present in the nucleotide and amino acid sequences of gp35 disclosed in the prior art. Indeed, the inventor has discovered that the prior art predicted amino acid sequence of gp35 lacks 77 amino acid residues at the N-terminus of the actual protein and that 15 of the 16 amino acid residues corresponding to the N-terminal residues of the prior art predicted gp35 are incorrect. The invention thus provides sequences of gp35 that correct these prior art errors.

The present invention thus relates to nucleotide sequences of *gp35* and amino acid sequences of encoded gp35 proteins, as well as derivatives and analogs thereof, and antibodies thereto.

As described by way of example *infra*, the present inventor has isolated and characterized the gene encoding bacteriophage T4 gp35, a tail component necessary for the formation of bacteriophage T4 tail fibers. The nucleotide sequence encoding gp35 was determined to be distinct from that previously reported in the NCBI database (Figure 3). According to the present invention, the *gp35* nucleotide sequence encodes a protein that has a different N-terminus and a molecular weight that is 24% greater than that predicted by the sequence in the NCBI database (nucleotides 4,127-5,011 of Figure 3).

In contrast to the prior art, by providing the correct sequence of gp35 (including the correct amino-terminal portion of the molecule), the present invention enables recombinant production and genetic manipulation of the gp35 protein.

5 In a preferred aspect, the present invention provides a purified bacteriophage gp35 protein that is not contained in a gel (e.g., a gel suitable in which to conduct electrophoresis).

In a specific embodiment, the invention relates to a composition comprising at least 1, 10, 50, 100 or 500 nanogram(s), 1, 10, 50, 100 or 500 microgram(s), or 1, 10, 50, 100 or 500 milligram(s), of purified non-denatured gp35 protein.

10 The *gp35* gene sequence of the invention can be a naturally occurring sequence or in variant form, whether natural, synthetic, or recombinant. In a specific embodiment, the gp35 protein is not native (i.e., not naturally occurring).

In a specific embodiment, the present invention relates to a bacteriophage T4 gp35 protein variant containing the amino acid sequence depicted in Figure 2 (SEQ ID NO:2)
15 wherein only conservative substitutions relative to the sequence in Figure 2 are made. The invention also relates to purified molecules comprising bacteriophage T4 gp35 protein fragments, which fragments consist of at least the amino acid sequence depicted in Figure 2 (SEQ ID NO:2) from amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79 or 81-93, as well as derivatives thereof, e.g., in which only conservative substitutions
20 relative to the sequence in Figure 2 are made. Nucleic acids encoding such proteins, and their complement, are also within the scope of the invention.

The invention additionally relates to proteins, derivatives, fragments or analogs containing an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85% or 90% identity to amino acids number 1 to 100 in Figure 2 over a 100 amino acid sequence. As
25 used herein, amino acid sequence homology refers to amino acid sequences having identical amino acid residues or amino acid sequences containing conservative changes in amino acid residues. In another embodiment, a gp35 homologous protein is one that shares the foregoing percentages of sequences identical with the naturally occurring gp35 protein over a 100 amino acid length.

30 The invention additionally relates to proteins, derivatives, fragments or analogs containing an amino acid sequence having at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85% or 90% identity to amino acids number 57 to 93 in Figure 2

over a 36 amino acid sequence. In another embodiment, a gp35 homologous protein is one that shares the foregoing percentages of sequences identical with the naturally occurring gp35 protein over a 36 amino acid length.

The invention also relates to proteins encoded by nucleic acids hybridizable to a
5 gp35 gene under non-stringent, moderately stringent, or stringent conditions. In a specific embodiment, such a protein is encoded by a nucleic acid hybridizable to a DNA having a nucleotide sequence consisting of the coding region of SEQ ID NO:1 or its complement.

As defined herein, a gp35 derivative may be a fragment or amino acid variant (*e.g.*, an insertion, substitution and/or deletion derivative) of the gp35 sequence shown in Figure
10 2. In a specific embodiment, such insertion, substitution and/or deletion occur outside of amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79 or 81-93 depicted in Figure 2.

The invention also relates to gp35 analogs.

The gp35 fragment, amino acid variant or analog of the invention is capable of
15 displaying one or more functional activities associated with a full-length native gp35 protein. Such functional activities include, but are not limited to, antigenicity, *i.e.*, the ability to bind to an anti-gp35 antibody, immunogenicity, *i.e.*, the ability to generate an antibody which is capable of binding a gp35 protein; the ability to bind (or compete with gp35 for binding) to a ligand for gp35; and the ability to multimerize with P36 and/or P34.
20 For an example of the latter, a functional ability of the gp35 protein is the ability of gp35 or a gp35-P36 oligomer to bind to P34 and/or the ability of gp35 to bind to P36.

In a specific embodiment, the invention provides gp35 fragments or variants that comprise at least a functionally active portion of the gp35 sequence shown in Figure 2 from amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79, or 81-93.

25 In a specific embodiment, the invention provides derivatives (including fragments) or analogs of a gp35 protein consisting of at least 8 contiguous amino acids, or of at least 15 contiguous amino acids, or of at least 20 contiguous amino acids, of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 24. In a preferred embodiment, this derivative or analog is able to be bound by an antibody directed
30 against a gp35 protein. In another preferred embodiment, the derivative or analog specifically binds the P34 homooligomer. Nucleic acids encoding such derivatives or analogs are also within the scope of the invention.

The invention further provides derivatives, fragments or analogs of a gp35 protein consisting of at least 40, 45, 50, 60, or 70 contiguous amino acid residues of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 100. In a specific embodiment, this gp35 derivative, fragment or analog lacks amino acid
5 residues 93 to 372.

The invention further relates to fragments (and derivatives and analogs thereof) of gp35 which comprise one or more functional domains of a gp35 protein, *e.g.*, the P36 or P34 binding domain, and/or retain the antigenicity of a gp35 protein (*i.e.*, are able to be bound by an anti-gp35 antibody). In specific embodiments, the fragments lack at least 10,
10 20, 30 or 40 contiguous amino acids of the sequence shown in Figure 2.

The invention also relates to gp35 proteins, derivatives and analogs in which internal peptide sequences are deleted without affecting the ability of gp35 to associate with its natural tail fiber partners P36 and/or P34. In a specific embodiment, the deletion occurs of contiguous amino acids selected from among amino acids 100-273. In other embodiments,
15 gp35 is modified so that it interacts only with other modified, and not native, tail fiber partners; exhibit thermolabile interactions with its partners; or contains, or is conjugated to, additional functional groups that enables it to interact with heterologous binding moieties.

The gp35 protein, or derivatives or analogs thereof, described herein, may be used for the production of anti-gp35 antibodies, which antibodies may be used in immunoassays
20 for the detection or measurement of gp35 protein.

The present invention also relates to a gp35 protein, derivative or analog that is modified in the domain that which forms an angle joint, to form an average angle that is different from the natural average angle of 137° ($\pm 7^{\circ}$) or 156° ($\pm 12^{\circ}$).

The present invention further relates to methods of production of the gp35 proteins,
25 derivatives and analogs, such as, for example, by recombinant means.

The present invention additionally provides for nanostructures comprising native or modified gp35 and native or modified bacteriophage tail fiber proteins. The nanostructures may be one-dimensional rods, two-dimensional polygons or open or closed sheets, or three-dimensional open cages or closed solids. The gp35 protein may be modified in
30 various ways to form novel structures with different properties for use as described in Section 5.8.

5.1. THE gp35 CODING SEQUENCES

gp35 DNA sequences and sequences complementary thereto are *gp35* nucleic acids provided by the present invention. Sequences hybridizable thereto, are also provided. Nucleic acids comprising *gp35* DNA or RNA sequences are also provided; in various
5 embodiments, at least 850, 880, 920, 960, or 1000 contiguous nucleotides of the *gp35* sequence in Figure 2, are in the nucleic acid. Also included within the scope of the present invention are nucleic acids comprising *gp35* DNA having the sequence depicted in Figure 2 (SEQ ID NO:2), or its corresponding RNA, which do not encode other bacteriophage T4 tail fiber proteins or functionally active portions thereof.

10 Nucleic acids can be single-stranded or double-stranded. In specific embodiments, isolated nucleic acids are provided that comprise at least 150, 175, 200, 225, 250, 275, or 285 contiguous nucleotides of nucleotides 1 to 285 in Figure 2.

In specific embodiments, the nucleic acids of the invention comprise the nucleotide sequences shown in Figure 2 that encode amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17,
15 57-93, 57-64, 66-79, or 81-93 of Figure 2.

In another embodiment, the nucleic acids comprise nucleotide numbers 1 to 1,116 of Figure 2.

The *gp35* nucleotide sequences of the invention, preferably do not contain in contiguous linkage sequences of a bacteriophage T4 genome that are naturally in contiguous
20 linkage flanking the *gp35* sequences (*i.e.*, 5' or 3' to the *gp35* gene). For example, the *gp35* nucleotide sequences can be contiguous with non-bacteriophage T4 nucleotide sequences of at least 10 nucleotides.

In a specific embodiment, the invention provides an isolated nucleic acid comprising a nucleotide sequence encoding a *gp35* protein having the amino acid sequence depicted in
25 Figure 2 (SEQ ID NO:2), operably linked to a heterologous promoter. By "heterologous promoter" is meant a promoter that is not the native T4 promoter that is operably linked to the *gp35* sequence in the bacteriophage T4 genome. In a specific embodiment, the promoter is not a bacteriophage T4 promoter. In a preferred embodiment, the nucleotide sequence encoding the *gp35* protein is that sequence depicted in Figure 2 (SEQ ID NO:1)
30 from nucleotide numbers 1 to 1,116 contiguous to a 3' termination codon.

In other specific embodiments, nucleic acids contain at least 850, 880, 920, 960, or 1000 contiguous nucleotides of a *gp35* DNA sequence operably linked to a promoter that is not a bacteriophage promoter (i.e., a heterologous promoter).

5 In a specific embodiment, the nucleic acid further comprises nucleotide sequences encoding other bacteriophage T4 proteins selected from the group consisting of *gp36* and *gp37*, and optionally the chaperon protein *gp57*, operably linked to the same or a different promoter. Preferably, native intergenic regions between the other bacteriophage T4 proteins are omitted.

The invention also provides single-stranded oligonucleotides for use as primers in
10 PCR that amplify a *gp35* gene or *gp35* sequence-containing fragment, e.g., an oligonucleotide having the sequence of a hybridizable portion (at least ~8 nucleotides) of *gp35*, and another oligonucleotide having the reverse complement of a downstream sequence in the same strand of *gp35*, such that each oligonucleotide primes synthesis in a direction toward the other. In one embodiment, the 5' oligonucleotide corresponds to
15 sequence flanking nucleotides 1-280 of Figure 2. In a specific embodiment, the 5' primer comprises a sequence upstream of nucleotide number 1 in Figure 2 and/or also comprises a nucleotide sequence shown in Figure 2 encoding an amino-terminal portion (i.e. at least the N-terminal amino acid) of *gp35*. In a specific embodiment, the oligonucleotide primers are preferably in the range of 10-35 nucleotides in length. A kit comprising in one or more
20 containers the foregoing primers is also provided.

The full length sequence for *gp35* is depicted in Figure 2 (SEQ ID NO:1), with the coding region thereof spanning nucleotide numbers 1 to 1,116. Sequence analysis of the nucleotide sequence of *gp35* of Figure 2 reveals an open reading frame of 1,116 nucleotides, encoding a protein of 372 amino acids (SEQ ID NO:2).

25 In accordance with the present invention, any polynucleotide sequence which encodes the amino acid sequence of a *gp35* product can be used to generate recombinant molecules which direct the expression of *gp35*. Included within the scope of the present invention are nucleic acids consisting of at least 8 nucleotides that are useful as probes or primers (i.e., a hybridizable portion) in the detection or amplification of *gp35*. In a
30 preferred embodiment, these probes or primers have a contiguous sequence contained in nucleotides 1 to 279 of Figure 2. The invention also relates to nucleic acid sequences hybridizable or complementary to the foregoing sequences or equivalent to the foregoing

sequences in that the equivalent nucleic acid sequences also encode a protein product displaying gp35 functional activity.

Nucleic acids encoding fragments and derivatives of gp35 are additionally described *infra*.

5 The invention also relates to nucleic acids hybridizable to or complementary to the above-described nucleic acids comprising *gp35* sequences. In specific aspects, nucleic acids are provided which comprise a sequence absolutely complementary to at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of a *gp35* gene, or, in particular, those portions encoding amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-
10 79, and/or 81-93 of Figure 2. In a specific embodiment, a nucleic acid which is hybridizable to a *gp35* nucleic acid, or to a nucleic acid encoding a *gp35* derivative, under conditions of low stringency is provided. By way of example and not limitation, procedures using such conditions of low stringency are as follows (see also Shilo and Weinberg, 1981, *Proc. Natl. Acad. Sci. USA* 78:6789-6792): Filters containing DNA are pretreated for 6 h at
15 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10⁶ cpm ³²P-labeled probe is used. Filters are incubated in
20 hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and reexposed to film. Other conditions of low stringency which may be
25 used are well known in the art (*e.g.*, as employed for cross-species hybridizations).

In another specific embodiment, a nucleic acid which is hybridizable to a *gp35* nucleic acid under conditions of high stringency is provided (see *infra*).

The DNA may be obtained by standard procedures known in the art from, for example, by chemical synthesis or by the cloning the DNA, or fragments thereof, purified
30 from a desired cell or phage. (See, for example, Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, *DNA Cloning: A Practical Approach*, MRL Press,

Ltd., Oxford, U.K. Vol. I, II.). Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

In the molecular cloning of the gene from DNA preparations, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at
5 specific sites using various restriction enzymes. The linear DNA fragments can then be separated according to size by standard techniques, including, but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography. See, for example, Innis et al., 1990, PCR protocols: A Guide to Methods and Applications, Academic Press, San Diego, California; Dieffenbach et al., 1995, PCR primer, A Laboratory Manual, Cold
10 Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, a *gp35* gene of the present invention or its specific RNA, or a fragment thereof, such as a probe or primer, may be isolated and labeled and then used in hybridization assays to detect
15 a generated *gp35* sequence (Benton, W. and Davis, R., 1977, *Science* 196:180; Grunstein, M., and Hogness, D., 1975, *Proc. Natl. Acad. Sci. USA* 72:3961). Those DNA fragments sharing substantial sequence homology to the probe will hybridize, e.g., under high stringency conditions. By way of example, the phrase "high stringency conditions" as used herein refers to those hybridizing conditions that (1) employ low ionic strength and high
20 temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50°C; (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075
25 M sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS.

It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known
30 restriction map. Further selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example,

DNA clones which hybrid-select the proper mRNAs, can be selected which produce a protein that has similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, binding activity or antigenic properties as known for gp35.

Alternatively, the gp35 protein may be identified by binding of labeled antibody to the
5 putatively gp35 expressing clones, *e.g.*, in an ELISA (enzyme-linked immunosorbent assay)-type procedure.

gp35 sequence can also be identified by mRNA selection by nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent
10 available, purified *gp35* DNA of a naturally occurring or modified *gp35* gene. Immunoprecipitation analysis or functional assays of the *in vitro* translation products of the isolated products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. Radiolabelled RNA or DNA may be used as a probe to identify the *gp35* DNA fragments from among other DNA
15 fragments.

Alternatives to isolating *gp35* DNA include, but are not limited to, chemically synthesizing the gene sequence itself from a known sequence. Other methods are known to those of skill in the art and are within the scope of the invention.

The identified and isolated DNA can then be inserted into an appropriate cloning
20 vector. A large number of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Such vectors include, but are not limited to, bacteriophages such as lambda or T4 derivatives, or plasmids such as PBR322 or pUC plasmid derivatives or the Bluescript vector (Stratagene). The insertion into a cloning
25 vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated
30 linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and *gp35* sequence may be modified by homopolymeric tailing. Recombinant molecules can be

introduced into host cells via transformation, transfection, infection, electroporation, or other methods known to those of skill in the art, so that many copies of the *gp35* sequence are generated.

5 In an alternative method, the desired *gp35* sequence may be identified and isolated after insertion into a suitable cloning vector in a "shot gun" approach. Enrichment for the desired DNA, for example, by size fractionization, can be done before insertion into the cloning vector.

10 In specific embodiments, host cells are transformed with recombinant DNA molecules that incorporate the isolated *gp35* sequence, or synthesized DNA sequence and enables generation of multiple copies of the sequence. Thus, the *gp35* sequence may be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted *gp35* sequence from the isolated recombinant DNA.

15 Oligonucleotides containing a portion of the *gp35* coding or non-coding sequences, or which encode a portion of the *gp35* protein (*e.g.*, primers for use in PCR) can be synthesized by standard methods commonly known in the art. Such oligonucleotides preferably have a size in the range of 8 to 25 nucleotides. In a specific embodiment herein, such oligonucleotides have a size in the range of 15 to 25 nucleotides or 15 to 35 nucleotides.

20 The *gp35* sequences provided by the instant invention include those nucleotide sequences encoding substantially the same amino acid sequences as found in native *gp35* proteins, and those encoded amino acid sequences with functionally equivalent amino acids, as well as those encoding other *gp35* derivatives or analogs, as described *infra* for *gp35* derivatives and analogs.

25

5.2. EXPRESSION OF *gp35* DNA

In accordance with the present invention, nucleotide sequences coding for a *gp35* protein, derivative (*e.g.* fragment) or analog thereof, can be inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription
30 and translation of the inserted protein-coding sequence, for the generation of recombinant DNA molecules that direct the expression of a *gp35* protein. Such *gp35* polynucleotide sequences, as well as other polynucleotides or their complements, may also be used in

nucleic acid hybridization assays, Southern and Northern blot analysis, etc. In a specific embodiment, a bacteriophage T4 *gp35* gene, or a sequence encoding a functionally active portion of a bacteriophage T4 *gp35* gene is expressed. In yet another embodiment, a derivative (e.g., fragment) of a bacteriophage T4 *gp35* gene is expressed.

5 Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent *gp35* amino acid sequence, is within the scope of the invention. Such DNA sequences include those which are capable of hybridizing to the *gp35* sequence of SEQ ID NO:1 under stringent conditions.

Altered DNA sequences which may be used in accordance with the invention
10 include deletions, additions or substitutions of different nucleotide residues resulting in a sequence that encodes the same or a functionally equivalent gene product. The gene product itself may contain deletions, additions or substitutions of amino acid residues within an *gp35* sequence, which result in a silent change, thus, producing a functionally equivalent *gp35* protein. Such amino acid substitutions may be made on the basis of
15 similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; amino acids with uncharged polar head groups having similar hydrophilicity values include the following: leucine, isoleucine, valine; glycine, alanine; asparagine,
20 glutamine; serine, threonine; phenylalanine, tyrosine.

The DNA sequences of the invention may be engineered in order to alter a *gp35* coding sequence for a variety of ends including, but not limited to, alterations which modify processing and expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert
25 new restriction sites, etc.

In an alternate embodiment of the invention, the coding sequence of *gp35* is synthesized in whole or in part, using chemical methods well known in the art. See, for example, Caruthers et al., 1980, *Nuc. Acids Res. Symp. Ser.* 7:215-233; Crea and Horn, 1980, *Nuc. Acids Res.* 9(10):2331; Matteucci and Caruthers, 1980, *Tetrahedron Letters*
30 21:719; and Chow and Kempe, 1981, *Nuc. Acids Res.* 9(12):2807-2817. Alternatively, the protein itself could be produced using chemical methods to synthesize a *gp35* amino acid sequence in whole or in part. For example, peptides can be synthesized by solid phase

techniques, cleaved from the resin, and purified by preparative high performance liquid chromatography (*e.g.*, *see* Creighton, 1983, Proteins Structures And Molecular Principles, W.H. Freeman and Co., N.Y. pp. 50-60). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure;
5 *see* Creighton, 1983, Proteins, Structures and Molecular Principles, W.H. Freeman and Co., N.Y., pp. 34-49).

In order to express a biologically active gp35 protein or functional equivalent thereof, a polynucleotide sequence encoding a gp35 protein, or derivative or analog thereof, is inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary
10 elements for the transcription and translation of the inserted coding sequence. The *gp35* gene products as well as host cells or cell lines transfected or transformed with recombinant gp35 expression vectors can be used for a variety of purposes. These include, but are not limited to, producing gp35 protein for use as an immunogen for generating antibodies (*i.e.*, monoclonal or polyclonal) that immunospecifically bind a gp35 protein and providing gp35
15 protein building blocks for nanostructures containing bacteriophage tail fiber proteins or protein derivatives.

5.2.1. EXPRESSION SYSTEMS

Methods known to those skilled in the art can be used to construct expression
20 vectors containing a gp35 coding sequence of interest (native, modified, or recombined) and appropriate transcriptional/translational control signals. These expression vectors typically contain selectable marker genes (usually conferring antibiotic resistance to transformed bacteria), sequences that allow replication of the plasmid to high copy number in *E. coli*, and a multiple cloning site immediately downstream of an inducible promoter and ribosome
25 binding site. Methods of constructing expression vectors containing a gp35 coding sequence include *in vitro* recombinant DNA techniques and synthetic techniques. *See*, for example, the techniques described in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory, N.Y. and Ausubel et al., 1989, Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley
30 Interscience, N.Y.

A variety of host-expression vector systems may be utilized to express a gp35 coding sequence. These systems are preferably bacteria transformed with recombinant

bacteriophage DNA or plasmid DNA expression vectors containing a gp35 coding sequence, but also include, but are not limited to, yeast transformed with recombinant yeast expression vectors containing an gp35 coding sequence; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing an gp35 coding
5 sequence; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing a gp35 coding sequence; or animal cell systems. The expression elements of these systems vary in their strength and specificities. Depending on the host/vector system utilized, any of a number of
10 suitable transcription and translation elements, including constitutive and inducible promoters, may be used in the expression vector. For example, when cloning in bacterial systems, inducible promoters such as P_l of bacteriophage λ , p_{trp}, p_{tac} (p_{trp}-lac hybrid promoter) and the like may be used. A preferred promoter is p_{lac} (with a *lac*ⁱ on the vector to reduce background expression). A second preferred promoter is pT7 ϕ 10, which is
15 specific to T7 RNA polymerase and is not recognized by *E. coli* RNA polymerase.

Examples of other host systems, include, but are not limited to; cloning in insect cell systems using promoters such as the baculovirus polyhedrin promoter; cloning in plant cell systems using promoters derived from the genome of plant cells (*e.g.*, heat shock promoters; the promoter for the small subunit of RUBISCO; the promoter for the chlorophyll a/b
20 binding protein) or from plant viruses (*e.g.*, the 35S RNA promoter of CaMV; the coat protein promoter of TMV); cloning in mammalian cell systems using promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5 K promoter); and generating cell lines that contain multiple copies of a gp35 DNA, SV40-, BPV- and EBV-
25 based vectors may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gp35 protein derivative or analog expressed. For example, when large quantities of gp35 protein, derivative or analog are to be produced for the generation of antibodies, vectors which direct the expression of high
30 levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited to, the *E. coli* expression vector pUR278 (Ruther et al., 1983, *EMBO J.* 2:1791), in which the gp35 coding sequence may be ligated into the vector in

frame with the *lacZ* coding region so that a hybrid AS-lacZ protein is produced; PIN vectors (Inouye & Inouye, 1985, *Nucleic acids Res.* 13:3101-3109; Van Heeke & Schuster, 1989, *J. Biol. Chem.* 264:5503-5509); and the like. PGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST) (Smith and Johnson, 1988, *Gene* 7:31-40). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The PGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety.

10 In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review *see*, Current Protocols in Molecular Biology, Vol. 2, 1988, Ed. Ausubel et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant et al., 1987, Expression and Secretion Vectors for Yeast, *in* Methods in Enzymology, Ed. Wu & Grossman, 1987, Acad. Press, N.Y. 153:516-544; Glover, 1986, DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3; and Bitter, 1987, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y. 152:673-684; and The Molecular Biology of the Yeast *Saccharomyces*, 1982, Eds. Strathern et al., Cold Spring Harbor Press, Vols. I and II.

In cases where plant expression vectors are used, the expression of a gp35 coding sequence may be driven by any of a number of promoters. For example, viral promoters such as the 35S RNA and 19S RNA promoters of CaMV (Brisson et al., 1984, *Nature* 310:511-514), or the coat protein promoter of TMV (Takamatsu et al., 1987, *EMBO J.* 6:307-311) may be used; alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi et al., 1984, *EMBO J.* 3:1671-1680; Broglie et al., 1984, *Science* 224:838-843); or heat shock promoters, *e.g.*, soybean hsp17.5-E or hsp17.3-B (Gurley et al., 1986, *Mol. Cell. Biol.* 6:559-565) may be used. These constructs can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, microinjection, electroporation, etc. For reviews of such techniques *see*, for example, Weissbach & Weissbach, 1988, *Methods for Plant Molecular Biology*, Academic Press, NY, Section VIII, pp. 421-463; and Grierson & Corey, 1988, *Plant Molecular Biology*, 2d Ed., Blackie, London, Ch. 7-9.

An alternative expression system which could be used to express a gp35 gene is an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. A gp35 coding sequence may be cloned into non-essential regions (for
5 example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example, the polyhedrin promoter). Successful insertion of a gp35 coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in
10 which the inserted gene is expressed. (*e.g.*, see Smith et al., 1983, J. Virol. 46:584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a gp35 coding sequence may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the
15 late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing gp35 in infected hosts. (*e.g.*, see Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Alternatively, the vaccinia 7.5 K promoter may be
20 used. (*See, e.g.*, Mackett et al., 1982, Proc. Natl. Acad. Sci. USA 79:7415-7419; Mackett et al., 1984, J. Virol. 49:857-864; Panicali et al., 1982, Proc. Natl. Acad. Sci. USA 79:4927-4931).

Other examples of commercially available vectors suitable for use in a bacteria host include, but are not limited to, the PET system (Novagen, Inc., Madison, WI) and
25 Superlinker vectors PSE280 and PSE380 (Invitrogen, San Diego, CA).

Specific initiation signals may also be required for efficient translation of an inserted gp35 coding sequence. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire *gp35* gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional
30 translational control signals may be needed. However, in cases where only a portion of a gp35 coding sequence is inserted, lacking the 5' end, exogenous translational control signals, including the ATG initiation codon, must be provided. Furthermore, the initiation

codon must be in phase with the reading frame of a gp35 coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer
5 elements, transcription terminators, etc. (see Bittner et al., 1987, *Methods in Enzymol.* 153:516-544).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications and processing (*e.g.*, cleavage) of protein products may be
10 important for the function of the protein. Different host cells have characteristic and specific mechanisms for post-transcriptional and post-translational processing and modification. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript,
15 and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, WI38, etc.

Preferred hosts for producing the proteins of the present invention are *E. coli* strains BL21 (DE3) and BL21 (DE/plys5) (NoVagen, Madison, Wisconsin).

For long-term, high-yield production of recombinant proteins, stable expression is
20 preferred. For example, cell lines which stably express a gp35 protein, derivative or analog may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with *gp35* DNA controlled by appropriate expression control elements (*e.g.*, bacterial promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), a selectable marker, and flanked by sequences that
25 promote homologous recombination. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows for the stable integration of the plasmid into host chromosomes. This method may advantageously be used to engineer bacterial strains
30 which express a gp35 protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler et al., 1977, *Cell* 11:223), hypoxanthine-guanine

phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc. Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy et al., 1980, Cell 22:817) genes can be employed in tk-, hgp^rt- or ap^rt- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler et al., 1980, Natl. Acad. Sci. USA 77:3567; O'Hare et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin et al., 1981, J. Mol. Biol. 150:1); and hyg^r, which confers resistance to hygromycin (Santerre et al., 1984, Gene 30:147). Recently, additional selectable genes have been described, namely trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, 1988, Proc. Natl. Acad. Sci. USA 85:8047); and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue, L., 1987, In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory, Ed.).

The present invention provides a method for producing a recombinant gp35 protein, derivative or analog comprising culturing a host cell transformed with a recombinant expression vector encoding a gp35 protein, derivative or analog, such that the gp35 protein, derivative or analog is expressed by the cell and recovering the expressed gp35 protein, derivative or analog.

5.2.2. IDENTIFICATION OF TRANSFECTANTS OR TRANSFORMANTS THAT EXPRESS gp35

The host cells which contain the coding sequence and which express the gp35 product or functionally active derivatives or analogs thereof may be identified by at least four general approaches; (a) DNA-DNA or DNA-RNA hybridization; (b) the presence or absence of "marker" gene functions; (c) assessing the level of transcription as measured by the expression of gp35 mRNA transcripts in the host cell; and (d) detection of the gene product as measured by immunoassay or by its biological activity.

In the first approach, the presence of the gp35 coding sequence inserted in the expression vector can be detected by DNA-DNA or DNA-RNA hybridization using probes

comprising nucleotide sequences that are homologous to the gp35 coding sequence, respectively, or derivatives (e.g., fragments) or analogs thereof.

In the second approach, the recombinant expression vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (e.g., resistance to antibiotics). For example, if the gp35 coding sequence is inserted within a marker gene sequence of the vector, recombinant cells containing the gp35 coding sequence can be identified by the absence of the marker gene function. Alternatively, a marker gene can be placed in tandem with a gp35 coding sequence under the control of the same or different promoter used to control the expression of the gp35 coding sequence. Expression of the marker in response to induction or selection indicates expression of the gp35 coding sequence.

In the third approach, transcriptional activity of gp35 can be assessed by hybridization assays. For example, RNA can be isolated and analyzed by Northern blot using a probe having sequence homology to a gp35 coding sequence or transcribed noncoding sequence or particular portions thereof. Alternatively, total nucleic acid of the host cell may be extracted and quantitatively assayed for hybridization to such probes.

In the fourth approach, the levels of a gp35 protein, derivative or analog product can be assessed immunologically, for example by Western blots, immunoassays such as radioimmuno-precipitation, enzyme-linked immunoassays and the like.

20

5.3. PURIFICATION OF THE EXPRESSED GENE PRODUCT

Once a recombinant which expresses the gp35 gene sequence is identified, the gene product can be analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labelling of the product followed by analysis by gel electrophoresis, immunoassay, or other detection methods known to those of skill in the art.

Once the gp35 protein is identified, it may be isolated and purified by standard methods including chromatography (e.g., ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. The functional properties may be evaluated using any suitable assay.

30

Alternatively, once a gp35 protein produced by a recombinant is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the protein can be synthesized by standard chemical methods known in the art (*e.g.*, see Hunkapiller et al., 1984, *Nature* 5 310:105-111).

In a specific embodiment, the invention relates to a purified gp35 protein that is not contained in a gel suitable for electrophoresis. In a preferred embodiment, the purified gp35 protein is not denatured.

In another specific embodiment, the invention relates to a composition containing at 10 least 1, 10, 50, 100 or 500 nanogram(s), 1, 10, 50, 100 or 500 microgram(s), or 1, 10, 50, 100 or 500 milligram(s), of purified non-denatured gp35 protein. In a preferred embodiment, this composition is not a gel suitable for electrophoresis.

In a specific embodiment of the present invention, such gp35 proteins, whether produced by recombinant DNA techniques or by chemical synthetic methods include, but 15 are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence substantially as depicted in Figure 2 (SEQ ID NO:2), as well as fragments and other derivatives, and analogs thereof.

5.4. GENERATION OF ANTIBODIES TO gp35

20 According to the invention, gp35 protein, its derivatives (*e.g.*, fragments), or analogs thereof, may be used as an immunogen to generate antibodies which recognize such an immunogen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library.

Various procedures known in the art may be used for the production of polyclonal 25 antibodies to a gp35 protein or derivative or analog.

In one embodiment, by way of example, gp35 specific antisera is prepared according to procedures as described by Edgar (1965, *Genetics* 52: 1187) and Ward (1970, *J. Mol. Biol.* 54:15). Briefly, whole T4 bacteriophage are used as an immunogen; the resulting antiserum is then adsorbed with tail-less phage particles, thus removing all antibodies 30 except those directed against the tail fiber proteins. In a subsequent step, different aliquots of the antiserum are adsorbed individually with extracts that each lack a particular tail fiber protein. For example, if an extract containing only tail fiber components gp34, gp36 and

gp37 (derived from a cell infected with a mutant T4 that does not produce gp35) is used for adsorption, the resulting antiserum will recognize only mature gp35 and dimerized gp35-P36 or gp35-P34. In an alternative embodiment, antibody is raised against purified tail fiber halves, *e.g.*, gp35-gp36-gp37. According to this embodiment, anti gp35-gp36-gp37 is
5 then adsorbed with gp36-gp37 to produce anti-gp35. In another embodiment, anti-gp35 is produced directly using purified gp35 proteins, derivatives or analogs thereof, as an immunogen. In another embodiment, monoclonal antibodies are generated against a gp35 protein sequence or analog thereof using techniques known in the art.

For the production of antibody, various host animals can be immunized by injection
10 with the native gp35 protein, or a synthetic version, derivative (*e.g.*, fragment) or analog thereof, including, but not limited to, rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including, but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions,
15 peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, BCG (bacille Calmette-Guerin) and corynebacterium parvum.

For preparation of monoclonal antibodies directed toward a gp35 protein sequence or analog thereof, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique
20 originally developed by Kohler and Milstein (1975, *Nature* 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, *Immunology Today* 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). In an additional embodiment of the invention, monoclonal antibodies can be
25 produced in germ-free animals utilizing recent technology (PCT/US90/02545).

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce gp35-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, *Science* 246:1275-1281) to
30 allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for gp35 proteins, derivatives, or analogs.

In one embodiment, a molecule comprising a fragment of the gp35 protein is used as an immunogen. In a preferred embodiment, the fragment used as the immunogen has a sequence that is all or a portion of amino acid residues 1 to 93, and lacks amino acid residues 94 to 373 in Figure 2. Since hydrophilic regions are believed most likely to
5 contain antigenic determinants, a peptide corresponding to or containing a hydrophilic portion of a gp35 protein is preferably used as immunogen.

Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include, but are not limited to: the F(ab')₂ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the F(ab')₂ fragment,
10 and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, *e.g.*, ELISA (enzyme-linked immunosorbent
15 assay). For example, to select antibodies which recognize a specific domain of a gp35 protein, one may assay generated hybridomas for a product which binds to a gp35 fragment containing such domain.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of the protein sequences of the invention, *e.g.*, for imaging these
20 proteins, measuring levels thereof, in diagnostic methods, etc.

A non-limiting method by which anti-gp35 may also be used to detect gp35 tail fiber proteins, derivatives or analogs, involves screening for bacterial colonies expressing proteins, derivatives or analogs by directly transferring the colonies, or, alternatively, samples of lysed or unlysed cultures, to nitrocellulose filters, lysing the bacterial cells on
25 the filter if necessary, and incubating with specific antibodies. Formation of immune complexes may then be detected by methods widely used in the art (*e.g.*, secondary antibody conjugated to a chromogenic enzyme or radiolabelled Staphylococcal Protein A.). This method is particularly useful to screen large numbers of colonies. In an alternative method, bacterial cells expressing the protein, derivative, or analog of interest are first
30 metabolically labelled with ³⁵S-methionine, followed by preparation of extracts and incubation with antiserum. The immune complexes may then be recovered by incubation with immobilized Protein A followed by centrifugation and resolution by SDS-PAGE.

5.5. STRUCTURE OF THE *gp35* GENE AND PROTEIN

The structure of the *gp35* gene and protein can be analyzed by any of various methods known in the art. Representative methods are set forth below.

5

5.5.1. GENETIC ANALYSIS

The cloned DNA corresponding to *gp35* can be analyzed by methods including, but not limited to, Southern hybridization (Southern, E.M., 1975, *J. Mol. Biol.* 98:503-517), Northern hybridization (*see, e.g.*, Freeman et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:4094-4098), restriction endonuclease mapping (Maniatis, T., 1982, *Molecular Cloning, A Laboratory*, Cold Spring Harbor, New York), and DNA sequence analysis. Polymerase chain reaction (PCR; U.S. Patent Nos. 4,683,202, 4,683,195, and 4,889,818; Gyllenstein et al., 1988, *Proc. Natl. Acad. Sci. USA* 85:7652-7656; Ochman et al., 1988, *Genetics* 120:621-623; Loh et al., 1989, *Science* 243:217-220) followed by Southern hybridization with a *gp35*-specific probe. Northern hybridization analysis can be used to determine the expression levels of *gp35*. The stringency of the hybridization conditions for both Southern and Northern hybridization, or dot blots, can be manipulated to ensure detection of nucleic acids with the desired degree of relatedness to the specific *gp35* probe used.

Restriction endonuclease mapping can be used to roughly determine the genetic structure of *gp35*. Restriction maps derived by restriction endonuclease cleavage can be confirmed by DNA sequence analysis.

DNA sequence analysis can be performed by any techniques known in the art, including, but not limited to, the method of Maxam and Gilbert (1980, *Meth. Enzymol.* 65:499-560), the Sanger dideoxy method (Sanger et al., 1977, *Proc. Natl. Acad. Sci. USA* 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699), or use of an automated DNA sequenator (*e.g.*, Applied Biosystems, Foster City, CA). The nucleotide sequence of a representative *gp35* gene comprises the sequence substantially as depicted in Figure 2 (SEQ ID NO:1), and described in Section 6, *infra*.

30

5.5.2. PROTEIN ANALYSIS

The amino acid sequence of a *gp35* protein, derivative, fragment or analog can be derived by deduction from the DNA sequence, or alternatively, by direct sequencing of the

protein, *e.g.*, with an automated amino acid sequencer. The amino acid sequence of a representative gp35 protein comprises the sequence substantially as depicted in Figure 2 (SEQ ID NO:2), and detailed in Section 6, *infra*, with the representative protein that is shown by amino acid numbers 1-372.

5 The gp35 protein sequence can be further characterized by a hydrophilicity analysis (Hopp, T., and Woods, K., 1981, *Proc. Natl. Acad. Sci. USA* 78:3824). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of the gp35 protein and the corresponding regions of the DNA sequence which encode such regions. Hydrophilic regions are predicted to be antigenic/immunogenic.

10 Secondary structural analysis (Chou, P., and Fasman, G., 1974, *Biochemistry* 13:222) can also be done, to identify regions of the gp35 protein that assume specific secondary structures.

 Manipulation, translation, and secondary structure prediction, as well as open reading frame prediction and plotting, can also be accomplished using computer software
15 programs available in the art.

 Other methods of structural analysis can also be employed. These include, but are not limited to, X-ray crystallography (Engstrom, A., 1974, *Biochem. Exp. Biol.* 11:7-13) and computer modeling (Fletterick, R., and Zoller, M. (eds.), 1986, *Computer Graphics and Molecular Modeling*, in *Current Communications in Molecular Biology*, Cold Spring
20 Harbor Laboratory, Cold Spring Harbor, New York).

5.6. gp35 PROTEINS, DERIVATIVES AND ANALOGS

 The invention further relates to gp35 proteins, derivatives (including, but not limited to, fragments) and analogs of gp35 proteins. Nucleic acids encoding gp35 proteins,
25 derivatives and analogs are also provided. Molecules comprising gp35 proteins, derivatives or analogs are also provided. In one embodiment, the gp35 proteins, derivatives or analogs are encoded by the gp35 nucleic acids described in Section 5.1 *supra*.

 The production and use of derivatives and analogs related to gp35 are within the scope of the present invention. In a specific embodiment, the derivative or analog is
30 functionally active, *i.e.*, capable of exhibiting one or more functional activities associated with a full-length, wild-type gp35 protein. As one example, such derivatives or analogs which have the desired immunogenicity or antigenicity can be used, for example, in

immunoassays, for inhibition of gp35 activity, etc. As another example, such derivatives or analogs which are able to bind bacteriophage T4 tail fiber proteins P36 and/or P34 are provided. Derivatives or analogs that retain a desired gp35 property of interest (*e.g.*, binding to tail fiber proteins), can be used as inhibitors of such property and its physiological correlates. A specific embodiment relates to a gp35 fragment that can be bound by an anti-gp35 antibody. Derivatives or analogs of gp35 can be tested for the desired activity by procedures known in the art, including, but not limited to, the assays described *infra*.

In particular, gp35 derivatives can be made by altering *gp35* sequences by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences which encode substantially the same amino acid sequence as a *gp35* gene may be used in the practice of the present invention. These include, but are not limited to, nucleotide sequences comprising all or portions of *gp35* which are altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the gp35 derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of a gp35 protein including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Conservative substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

In one embodiment, the invention relates to non-native bacteriophage T4 gp35 proteins, derivatives or analogs in which only conservative substitutions relative to the sequence in Figure 2 are made.

The invention also relates to non-native molecules encoded by a nucleic acid that is capable of hybridizing to gp35 coding sequence (SEQ ID NO:1), under stringent, moderately stringent, or nonstringent conditions.

5 In another embodiment, the invention relates to proteins, derivatives or analogs comprising the amino acid sequence depicted in Figure 2 (SEQ ID NO:2) from amino acid residues 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-93, 57-64, 66-79, or 81-93. In another embodiment, these proteins contain only conservative substitutions relative to the sequence in Figure 2.

10 The invention additionally relates to proteins, derivatives or analogs, comprising an amino acid sequence that has at least 60%, 65%, 70%, 75%, 80%, 85%, or 90% amino acid sequence homology, to bacteriophage T4 gp35 amino acids number 1 to 100 in Figure 2 over a 100 amino acid sequence.

The invention further relates to proteins, derivatives, fragments or analogs comprising an amino acid sequence sharing at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 15 65%, 70%, 75%, 80%, 85% or 90% homology to amino acids numbers 57 to 93 in Figure 2 over a 36 amino acid sequence.

The invention further provides derivatives, fragments or analogs of a gp35 protein consisting of at least 8, 15, or 20 contiguous amino acids of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 24. In one 20 embodiment, the derivative, fragment or analog is not native and contains only conservative substitutions relative to the sequence in Figure 2. In a preferred embodiment, the derivative or analog additionally displays one or more functional activities of a gp35 protein. In another preferred embodiment, the derivative, fragment or analog specifically binds P34 and/or P36. In another preferred embodiment, the derivative or analog is able to be bound 25 by an antibody directed against a gp35 protein in which only conservative substitutions relative to the sequence in Figure 2 are made.

The invention also provides derivatives or analogs of a gp35 protein consisting of at least 40, 45, 50, 60, or 70 contiguous amino acid residues of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 100. In a specific 30 embodiment, this derivative lacks amino acid residues 93 to 372.

Tail fiber assembly takes place in a predetermined, ordered interaction of specific bacteriophage protein subunits. The angled joint of the tail fiber is formed by the two step

process in which first, the N-terminus of P36 attaches to the carboxy terminal region of a gp35 monomer and second the N-terminal region of gp35-P36 oligomer then attaches to the C-terminus of P34. In a specific embodiment, a gp35 mutant/derivative or analog is provided in which the interaction of the gp35 derivative or analog with P34 is independent
5 of the gp35 first interacting with P36.

In another embodiment of the invention, gp35 derivatives or analogs form average angles with other tail fiber proteins that are different from the native angle of 137° or 158°. In specific embodiments, the angle joint forms average angles of less than about 90°, 100°, 110°, 120°, or 125°, or more than about 145°, 155°, 165°, under conditions wherein the
10 wild-type gp35 protein forms an angle of 137° when combined with P36-P37 and P34 dimers or trimers. In other embodiments, the angle joint of gp35 proteins, derivatives or analogs exhibit more or less flexibility than the native polypeptide. gp35 sequence variants can be screened for the ability to form such an angle.

Thermolabile structures have many uses in nanostructure construction, such as, for
15 example, initiation of structure assembly at low temperature and subsequent inactivation of and separation from the initiator at high temperature. In one embodiment of the invention, gp35 derivatives and analogs exhibit thermolabile interactions with cognate partners. For example, in one embodiment the interaction of a gp35 derivative with a P36 protein oligomer of bacteriophage T4 is unstable at a temperature of about 40°C, 45°C, 50°C,
20 55°C or 60°C (see Section 7). In another embodiment, the interaction of a gp35 derivative with a P34 protein oligomer of bacteriophage T4 is unstable at a temperature of about 40°C, 45°C, 50°C, 55°C or 60°C (see Section 7). In a specific embodiment, the thermolabile interaction between gp35 and cognate partners is reversible, thereby permitting reattachment of the appropriate termini when the lower temperature is restored, in another
25 specific embodiment, this interaction is irreversible.

In another specific embodiment, the gp35 derivative or analog interacts with only mutant cognate partners (e.g., see Section 7).

In another embodiment, gp35 derivatives or analogs contain a mutant amino acid sequence, or are conjugated to a fixed group, that confers specific binding properties on the
30 entire molecule, e.g., sequences derived from avidin that recognize biotin, sequences derived from immunoglobulin heavy chain that recognize Staphylococcal A protein, sequences derived from the Fab portion of the heavy chain of monoclonal antibodies to

which their respective Fab light chain counterparts could attach and form an antigen-binding site, immunoactive sequences that recognize specific antibodies, or sequences that bind specific metal ions (e.g., divalent metal ions). These ligands may be immobilized to facilitate purification and/or assembly.

5 In a specific embodiment of the invention, proteins consisting of or comprising a fragment of a gp35 protein consisting of at least 8 (continuous) amino acids of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids number 1 to 24. In other embodiments, the fragment consists of at least 15 or 20 amino acids of the gp35 protein depicted in Figure 2 from amino acids number 1-24.

10 The invention also provides fragments of a gp35 protein consisting of at least 40, 45, 50, 55, 60, or 70 contiguous amino acid residues of the gp35 sequence in Figure 2 (SEQ ID NO:2) from amino acids 1-100. In specific embodiments, such fragments are not larger than 75, 100 or 150 amino acids. In other specific embodiments, such fragments lack amino acid number 93 to 372 in Figure 2. Derivatives or analogs of gp35 include, but are not
15 limited to, those molecules comprising regions that are substantially homologous to gp35 or fragments thereof (e.g., in various embodiments, at least 60% or 70% or 80% or 90% or 95% identity over an amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art) or whose encoding nucleic acid is capable of hybridizing to a gp35 coding sequence,
20 under stringent, moderately stringent, or nonstringent conditions.

The gp35 derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned gp35 sequence can be modified by any of numerous strategies known in the art (Maniatis, T., 1990, *Molecular Cloning, A Laboratory*
25 *Manual*, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. In the production of the gene encoding a derivative or analog of gp35, care should be taken to ensure that the modified gene remains within the same translational reading frame as gp35, uninterrupted
30 by translational stop signals, in the gene region where the desired gp35 activity is encoded.

Additionally, the gp35-encoding nucleic acid sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to

create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, including, but not limited to, chemical mutagenesis, *in vitro* site-directed mutagenesis (Hutchinson, C., et al., 1978, *J. Biol. Chem* 253:6551), PCR
5 amplification using primers with altered sequences, etc.

Manipulations of the gp35 sequence may also be made at the protein level. Included within the scope of the invention are gp35 protein fragments or other derivatives or analogs which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking
10 groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including, but not limited to, specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH₄; acetylation, formylation, oxidation, reduction; etc.

15 In addition, analogs and derivatives of gp35 can be chemically synthesized. For example, a peptide corresponding to a specific portion of a gp35 protein (see Section 5.6.1), or which mediates the desired activity *in vitro*, can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the gp35 sequence. Non-
20 classical amino acids include, but are not limited to, the D-isomers of the common amino acids, α -amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, γ -Abu, ϵ -Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β -alanine, fluoro-amino
25 acids, designer amino acids such as β -methyl amino acids, C α -methyl amino acids, N α -methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

In another embodiment, the gp35 derivative is a molecule comprising a region of homology with a gp35 protein. By way of example, in various embodiments, a first protein
30 region can be considered "homologous" to a second protein region when the amino acid sequence of the first region is at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, or 95% identical, when compared to any sequence in the second region of an equal number of

amino acids as the number contained in the first region or when compared to an aligned sequence of the second region that has been aligned by a computer homology program known in the art. For example, a molecule can comprise one or more regions homologous to a gp35 region (see Section 5.6.1) or a full-length gp35 protein.

5 In another embodiment, the gp35 proteins, derivatives, fragments or analogs of the invention are combined with other tail fiber proteins, derivatives, fragments and/or analogs, to form polygons. In a preferred embodiment, a polygon is formed using the gp35 protein, derivative, or analog of the invention in combination with a P36-34 chimera rod unit as described in PCT Publication WO 96/11947, dated April 25, 1996.

10

5.7. ASSAYS OF gp35 PROTEINS, DERIVATIVES AND ANALOGS

The functional activity of gp35 proteins, derivatives and analogs can be assayed by various methods.

15

For example, in one embodiment, where one is assaying for the ability to bind or compete with wild-type gp35 for binding to anti-gp35 antibody, various immunoassays known in the art can be used, including, but not limited to, competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel
20 diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In one embodiment, antibody binding is detected by
25 detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labelled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

30

In another embodiment, where a gp35-binding protein is identified (*e.g.*, P34 and P36), the binding can be assayed, by means well-known in the art.

A nonlimiting method by which antibodies specific to gp35 proteins may be used to assay for the ability of gp35 proteins, derivatives or analogs to associate with other tail fiber proteins involves screening for bacterial colonies expressing mature tail fiber proteins by directly transferring the colonies, or, alternatively, samples of lysed or unlysed cultures, to
5 nitrocellulose filters, lysing the bacterial cells on the filter if necessary, and incubating with antibodies specific for gp35 and its binding partner and detecting the formation of immune complexes by methods widely used in the art (e.g., secondary antibody conjugated to a chromogenic enzyme or radiolabelled Staphylococcal Protein A). Another nonlimiting method involves metabolically labelling bacterial cells expressing gp35 with ³⁵S-
10 methionine, preparing and incubating extracts of these cells with gp35 antiserum, recovering immune complexes by incubation with immobilized Protein A followed by centrifugation, and resolving the proteins by SDS-polyacrylamide gel electrophoresis.

A nonlimiting competitive assay for testing whether gp35 derivatives or analogs such as internally deleted tail fiber proteins that do not permit phage infection nonetheless
15 retain the ability to associate with their appropriate partners utilizes an *in vitro*, complementation system which involves mixing a bacterial extract containing the modified gp35 tail fiber protein with a second extract prepared from cells infected with a phage that is a *gp35* null mutant and therefore does not produce gp35. After several hours of incubation, a third extract is added that contains wild-type gp35, and incubation is continued for several
20 additional hours. Finally, the extract is titered for infectious phage particles by infecting *E. coli* and quantifying the phage plaques that result. A modified gp35 protein, derivative or analog that correctly associates with its tail fiber partners is incorporated into tail fibers in a non-functional manner in the first mixture, thereby preventing the incorporation of the wild-type version of the protein after addition of the third extract; the result is a reduction in
25 the titer of the resulting phage sample. By contrast, if the modified gp35 protein, derivative or analog is unable to associate with its binding partner, it will not be incorporated into phage particles in the first mixture and, thus, will not compete with assembly of intact phage particles when the third extract is added; the phage titer should thus be equivalent to that observed when no modified gp35 is added in to the first mixture (a negative control).

30 Assays for testing whether gp35 proteins, derivatives, such as internally deleted proteins, or analogs that do not permit phage infection nonetheless retain the ability to associate with appropriate tail fiber partners can also be performed *in vivo*. These assays

detect the ability of gp35 proteins, derivatives, or analogs to compete with normal phage parts for assembly, thus reducing the burst size of a wild-type phage infecting the same host cell in which gp35 proteins, derivatives, or analogs are recombinantly expressed. Thus, expression from an expression vector encoding the gp35 proteins, derivative, or analogs is
5 induced inside a cell, which cell is then infected by a wild-type phage. Inhibition of wild-type phage production demonstrates the ability of the recombinant gp35 protein, derivative, or analog to associate with the appropriate tail fiber proteins of the phage.

The above-described methods may be used, alone and in combination, in the design and production of different types of modified gp35 tail fiber proteins. For example, a
10 preliminary screen of a large number of bacterial colonies for those expressing properly associated P34-gp35 and/or gp35-P36 complexes will identify positive colonies, which can then be individually tested by *in vitro* complementation.

Other methods will be known to the skilled artisan and are within the scope of the invention.

15

5.8. APPLICATIONS OF NANOMETER STRUCTURE

The gp35 proteins, derivatives, and analogs of the invention have use in the construction of nanostructures. The uses of such nanostructures are manifold and include applications that require highly regular, well-defined arrays of fibers, cages, or solids, which
20 may include specific attachment sites that allow them to associate with other materials.

In one embodiment, a three-dimensional hexagonal array of tubes is used as a molecular sieve or filter, providing regular vertical pores of precise diameter for selective separation of particles by size. Such filters can be used for sterilization of solutions (*i.e.*, to remove microorganisms or viruses), or as a series of molecular-weight cut-off filters. In
25 this case, the protein components of the pores may be modified so as to provide specific surface properties (*i.e.*, hydrophilicity or hydrophobicity, ability to bind specific ligands, etc.). Among the advantages of this type of filtration device is the uniformity and linearity of pores and the high pore to matrix ratio.

In another embodiment, long one-dimensional fibers are incorporated, for example,
30 into paper or cement or plastic during manufacture to provide added wet and dry tensile strength.

In a further embodiment, different nanostructure arrays are impregnated into paper and fabric as anti-counterfeiting markers. In this case, a simple color-linked antibody reaction (such as those commercially available in kits) is used to verify the origin of the material. Alternatively, such nanostructure arrays could bind dyes or other substances,
 5 either before or after incorporation to color the paper or fabrics or modify their appearance or properties in other ways.

It will be apparent to one skilled in the art that the nanostructures comprising recombinant gp35 and its derivatives, fragments and analogs include, but are not limited to, other polygonal structures such as octagons, as well as open solids such as tetrahedrons and
 10 icosahedrons formed from triangles and boxes formed from squares and rectangles. The range of structures is limited only by the types of angle units and the substituents that can be engineered on the different axes of the rod units. For example, other naturally occurring angles are found in the fibers of bacteriophage T7, which has a 90° angle (Steven et al., *J. Mol. Biol.* 200: 352-365, 1988).

15 The use of bacteriophage tail fiber components in the construction of nanostructures is further described in PCT Publication WO 96/11947, dated April 25, 1996, which is incorporated by reference herein in its entirety.

Additionally, the gp35 proteins, derivatives, fragments and analogs of the invention have use in the study and research of the bacteriophage T4 life cycle.

20

5.9. NANOMETER STRUCTURE FORMATION

Bacteriophage T4 tail fiber proteins gp34, gp35, gp36, and gp37 are produced naturally following infection of *E. coli* cells by intact T4 phage particles. The structure of the T4 bacteriophage tail fiber (illustrated in Figure 1) can be represented schematically as
 25 follows (N= amino terminus, C= carboxy terminus): N[P34]C - N[gp35]C - N[P36]C - N[P37]C. P34, P36, and P37 homooligomers are stiff and rod-shaped proteins in which two identical β sheets, oriented in the same direction, are fused face-to-face by hydrophobic interactions between the sheets juxtaposed with a 180° rotational axis of symmetry through the long axis of the rod. gp35, by contrast, is a monomeric polypeptide that attaches
 30 specifically to the N-terminus of a P36 homooligomer and then to the C-terminus of a P34 homooligomer and forms an angle joint between two rods at an average angle of $137^\circ (\pm 7^\circ)$ or $156^\circ (\pm 12^\circ)$.

During T4 infection of *E. coli*, gp37 (the monomeric 109 Kda translation product of gene 37) forms the homooligomer P37, with the aid of 2 accessory (chaperon) proteins, gp57 and gp38; this process is believed to initiate near the C-terminus of gp37. Once P37 is formed, the N-terminus of P37 initiates the oligomerization of two gp36 molecules of 23 Kda each, in a butt-end joint to form the P36 homooligomer rod. The N-terminus of P36 then attaches to the carboxy terminal region of a gp35 monomer; this interaction stabilizes P36 and forms the flexible angle joint of the tail fiber. The amino terminal region of gp35 then attaches to the C-terminus of P34 (the homooligomerization of which requires the chaperon protein gp57). This regulation of self assembly of the tail fiber by a predetermined, ordered interaction of specific subunits results in the production of a structure of exact specifications from a random mixture of the tail fiber subunit components. Thus, self assembly of the tail fiber is regulated by a predetermined, ordered interaction between specific subunits whereby structural maturation caused by formation of the first subassembly permits interaction with new (previously disallowed) subunits. This results in the production of a structure of exact specifications from a random mixture of the components.

In one embodiment, the nanostructures of the invention are composed of tail fiber chimers, such as for example, P36-34, which is an oligomer of the fusion protein gp36-34; gp36-34 consists of a portion of gp36 containing the amino terminus fused to a portion of gp34 containing the carboxy terminus. Expression vectors encoding such chimers may be constructed using recombinant technology known in the art. Such chimers have novel functional properties, including but not limited to rod domains and/or N- and C-termini combinations that are different from native tail fiber proteins. Chimers having novel N- and C- termini combinations allow for new patterns for joining different rod segments. For example, polygon nanostructures may be generated using P36-34 chimeric fusion proteins and gp35. The creation of constructs encoding tail fiber fusion chimers, such as P36-34, and their use in generating nanostructures, is further described in PCT Publication WO 96/11947, dated April 25, 1996, which is incorporated by reference herein in its entirety.

Recombinant expression of the proteins of the present invention in *E. coli* as described above results in the synthesis of large quantities of protein, and allows the simultaneous expression and assembly of different components in the same cells. The methods for scale-up of recombinant protein production are straightforward and widely

known in the art, and many standard protocols can be used to recover native and modified tail fiber proteins from a bacterial culture.

In a preferred embodiment, recombinant gp35 is isolated for use by growing host cells transformed or adsorbed with nucleotide sequence encoding a gp35 protein having the amino acid sequence depicted in Figure 2, operably linked to a heterologous promoter, under conditions in which the gp35 encoding nucleic acid is expressed, and isolating gp35 from the resulting culture by standard methods.

P34, P36-P37, P37 and chimers derived therefrom, such as for example, P36-34, are purified from phage-infected (or recombinant) *E. coli* cultures as mature oligomers. gp35 protein, derivatives or analogs thereof are purified as monomers. Standard methods may be utilized to isolate and purify the nanostructure components, these methods include but are not limited to: chromatography on molecular sieve, ion-exchange, and/or hydrophobic matrices; preparative ultracentrifugation; and affinity chromatography, using as the immobilized ligand specific antibodies or other specific binding. For example, if the proteins have been engineered to include heterologous domains that act as ligands or binding sites, the cognate partner may be immobilized on a solid matrix and used in affinity purification. For example, such a heterologous domain can be avidin, which binds to a biotin-coated solid phase.

In an alternative preferred embodiment, several phage tail fiber components, and where necessary, chaperon proteins such as gp57 and gp37 required for homooligomerization, are co-expressed in the same bacterial cells, and sub-assemblies of larger nanostructures are purified subsequent to limited *in vivo* assembly, using the methods enumerated above.

In one embodiment, the purified nanostructure components and/or subassemblies are combined *in vitro* under conditions where assembly of the desired nanostructure occurs at temperatures between about 4°C and about 37°C, and at pH's between about 5 and about 9. For a given nanostructure, optimal conditions for assembly (*i.e.*, type and concentration of salts and metal ions) are easily determined by routine experimentation, such as by changing each variable individually and monitoring formation of the appropriate products.

In an alternate embodiment, one or more crude bacterial extracts are prepared, mixed, and assembly reactions are allowed to proceed prior to purification.

In some cases, one or more purified components assemble spontaneously into the desired structure, without the necessity for initiators. In other cases, an initiator is required to nucleate the polymerization of the nanostructure. This offers the advantage of localizing the assembly process (*i.e.*, if the initiator is immobilized or otherwise localized) and of
5 regulating the dimensions of the final structure. For example, rod components that contain a functional P36 homooligomer C-terminus require a functional P37 homooligomer N-terminus to initiate rod formation stoichiometrically; thus, altering the relative amount of initiator and rod component will influence the average length of rod polymer. If the ratio is n , the average rod will be approximately (P37 - P36) n --N-terminus P37-P37 C-terminus.

10 In still other cases, the final nanostructure is composed of two or more components that cannot self-assemble individually, but only in combination with each other. In this situation, alternating cycles of assembly can be staged to produce final products of precisely defined structure.

In one embodiment, polygons are assembled using gp35 and P36-34 chimera.

15 According to this embodiment, gp57 is used to chaperon the homodimerization of gp36-34 to P36-34. P36-34 chimera is added to a solution containing a gp35 initiator that optionally is reversibly immobilized using methods known in the art, so as to allow binding of P36-34 chimera. According to this embodiment, gp35 and P36-34 are administered as a mixture or sequentially to form the desired polygon structure. The type of polygon that is formed
20 using this protocol depends upon the length of rod units and the angle formed by the angle joint. For example, alternating rod units of different sizes can be used. In addition, variant gp35 polypeptides that form angles different than the natural angle can be used, allowing the formation of different regular polygons. Furthermore, for a given polygon with an even number of sides and equal angles, the sides in either half can be of any size provided the
25 two halves are symmetric. The creation of constructs encoding tail fiber fusion chimeras, such as P36-34, and their use in generating polygon nanostructures, is further described in PCT Publication WO 96/11947, dated April 25, 1996, which is incorporated by reference herein in its entirety.

When an immobilized initiator is used, it may be desirable to remove the
30 polymerized unit from the matrix after staged assembly. For this purpose specialized initiators are engineered so that the interaction with the first rod component is rendered reversibly thermolabile. For example, where a nanostructure is assembled that is attached

to a solid matrix via gp34, one way in which to detach the nanostructure to bring it into solution is to use a mutant (thermolabile) gp34 that can be made to detach upon exposure to a higher temperature (*e.g.*, 40°C). Such a mutant gp34, termed T4 tsB45, having a mutation at its C-terminal end such that gp34 attaches to the distal tail fiber half at 30°C, but can be
5 separated from it *in vitro* by incubation at 40°C in the presence of 1% SDS (unlike wild-type T4 which are stable under these conditions), has been reported (Seed, 1980, Studies of the Bacteriophage T4 Proximal Half Tail Fiber, Ph.D. Thesis, California Institute of Technology), and can be used. Using a reversibly thermolabile matrix band nanostructure/component, the polymer can be easily separated from the matrix-bound
10 initiator, thereby permitting: easy preparation of stock solutions of uniform parts or subassemblies, and re-use of the matrix-bound initiator for multiple cycles of polymer initiation, growth, and release.

The following examples are intended to illustrate the present invention without limiting its scope.

15

6. EXAMPLE: CLONING AND CHARACTERIZATION OF THE BACTERIOPHAGE *gp35* GENE

As described herein, the present inventor has isolated and characterized the T4 bacteriophage *gene 35*, a gene encoding a tail fiber protein which functions to join the
20 rodlike proximal and distal halves of the bacteriophage tail fibers. Phage T4 *gp35* is located between *gene 34* and *gene 36*. A sequence for *gp35* is available on the NCBI database (NCBI.NIH.GOV) within the sequence T4g34-t (nucleotides 4188-5075; see Figure 3). The NCBI sequence predicts that the *gp35* open reading frame, ORF35 encodes a putative protein having a molecular weight of 32,334 Daltons. However, the present inventor
25 noticed that this deduced molecular weight was discrepant with a reported molecular weight of *gp35* as determined by SDS-PAGE of 39,000-40,000 Daltons ("The T4 Book": Molecular Biology of Bacteriophage T4 (1994, Jim Karam editor, ASM Press, Wash. DC, pg. 507 and pg. 514). In addition, the NCBI database predicts a 241 nucleotide open reading frame, ORF34.1, located between *gene 34* and *gene 35* which encodes a protein
30 having a predicted molecular weight of 7,334 Daltons (in a different reading frame from ORF35). The inventor predicted that the NCBI sequence of *gp35* was incorrect and that the two open reading frames, ORFS 34.1 and 35 are actually connected to form a single ORF35

encoding a protein of about 40,000 Daltons. According to this postulation ORF34.1 encodes the N-terminus of gp35.

To prove this hypothesis, the inventor cloned his postulated *gp35* open reading frame by polymerase chain reaction (PCR) of the phage DNA between the 5'-ATG start
5 codon of ORF34.1 and the 3'-TAA stop codon of ORF35, a sequence of approximately 1,120 nucleotides in length, into an inducible expression plasmid, pT7-5, having appropriately situated RNA polymerase and ribosome binding sites and a *lacZ* promoter. Upon induction of expression of the insert from the *lacZ* promoter with IPTG, only one new heavy band (relative to uninduced cells) was apparent on SDS-PAGE, at 41,000 Daltons.
10 There was no visible band at either 32,000 Daltons or 7,000 Daltons.

Sequence analysis of the PCR generated insert revealed that *gp35* contains a single ORF of 1,119 nucleotide pairs having 373 codons, of which 372 encode a protein having a putative molecular weight of 40,096 Daltons. The terminal codon of the *gp35* open reading frame is the ochre stop codon, TAA. This 1,119 nucleotide sequence was compared with
15 the 1,121 nucleotide sequence from the NCBI database using the FASTA program. Six differences were detected between the sequence and that of the NCBI sequence. These six differences are: deletion of the adenine at nucleotide 22 of the NCBI sequence; insertion of a thymine between the adenine at nucleotide 49 and the thymine at nucleotide 50 of the NCBI sequence; deletion of the cytosine at nucleotide 170 of the NCBI sequence; change of
20 nucleotide 238 from a thymine to a cytosine of the NCBI sequence; deletion of the thymine at nucleotide 280 of the NCBI sequence; and change of nucleotide 557 of the NCBI sequence from an adenine to a guanine.

The sequence of the N-terminal 10 residues of the induced protein generated from the expression vector construct were determined to be identical to the first ten residues the
25 inventor predicted for the new *gp35* ORF. The determination of residues 8, 9 and 10 in the induced protein to be phenylalanine, glycine and glutamine, instead of the isoleucine, tryptophan and threonine residues respectively predicted for ORF 34.1 of the NCBI database sequence proves that the new *gp35* ORF sequence is correct and that the adenine located at nucleotide 22 in the NCBI sequence, is the result of a sequencing error and is not
30 actually present in bacteriophage T4 *gene 35*. The inventor has therefore shown that the correct *gp35* sequence is not that previously reported, but actually is a larger protein with a different N-terminus, that is 24% heavier than that predicted from the published sequence.

The correct *gp35* sequence encodes 77 more N-terminal amino acid residues than the NCBI sequence. Additionally, 15 of the first 16 N-terminal residues encoded by the NCBI sequence are incorrect.

Nucleic acid and protein database analysis of the new *gene 35* sequence and its
5 encoded product fails to reveal significant homology with other sequences in the databases.

7. **EXAMPLE: ISOLATION OF THERMOLABILE PROTEINS FOR SELF-ASSEMBLY**

A variant (temperature-sensitive) *gp35* that permits heat induced separation of the
10 *gp35*-P36 junction may be formed by mutagenizing the 3' region of *gp35* DNA (encoding the carboxy terminal region of *gp35*) with randomly doped oligonucleotides. Randomly doped oligonucleotides are prepared during chemical synthesis of oligonucleotides, by adding a trace amount (up to a few percent) of the other three nucleotides at a given position, so that the resulting oligonucleotide mix has a small percentage of incorrect
15 nucleotides at that position (Hutchison et al., 1991, Methods Enzymol. 202:356). The mutagenized DNA fragment is then recombined into T4 phage by infection of the cell containing the mutagenized DNA by a T4 phage containing two amber mutations flanking the mutagenized region. Following a low-multiplicity infection, non-amber phage are selected at low temperature on *E. coli* Su^o at 30°C. The progeny of these plaques are
20 resuspended in a buffered solution and challenged by heating at 60°C. At this temperature, wild-type tail fibers remain intact and functional, whereas the thermolabile versions release the P36 units and thus render those phage non-infectious.

At this stage, wild type phage are removed either by adsorbing the wild type phage to sensitive bacteria and sedimenting (or filtering out) the bacteria with the adsorbed wild
25 type phage or by reacting the lysate with anti-*gp35*-P36 specific antibody, followed by immobilized Protein A and removal of adsorbed wild type phage. Either of these methods leaves the noninfectious mutant phage particles in the supernatant fluid or filtrate, from which they can be recovered. The non-infectious phage lacking terminal *gp35*-P36 moieties are then urea treated with 6M urea, and mixed with bacterial spheroplasts to permit
30 infection at low multiplicity whereupon they replicate at low temperature and release progeny. Alternatively, infectious phage are reconstituted by *in vitro* incubation of the mutant phage with wild type P36 at 30°C; this is followed by infection of intact bacterial

cells using a standard protocol. The latter method of infection specifically selects mutant phage in which the thermolability of the gp35-P36 junction is reversible.

Using either method of infection, the phage populations are subjected to multiple rounds of selection, after which individual phage particles are isolated by plaque
5 purification at 30°C. Finally, the putative mutants are evaluated individually for: loss of infectivity after incubation at high temperatures (40-60°C), as measured by a decrease in titer; loss of P36 after incubation at high temperature, as measured by decrease in binding of gp35-P36-specific antibody to phage particles; and morphological changes in the tail fibers after incubation at high temperatures, as assessed by electron microscopy.

10 After the mutants are isolated and their phenotypes confirmed, the *gp35* gene is preferably sequenced. If the mutations localize to particular regions or residues, those sequences are preferably targeted for site-directed mutagenesis to optimize the desired characteristics.

Subsequently, the mutant *gene 35* is cloned into an expression plasmid and
15 expressed individually in *E. coli*. The mutant gp35 protein is then purified from bacterial extracts and used in *vitro* assembly reactions.

In a similar fashion, gp35 variants can be isolated that exhibit a thermolabile interaction with P34. In contrast to the localized mutagenesis described above, the screen for gp35 mutants exhibiting a thermolabile interaction with P34 involves random doped
20 oligonucleotide mutagenesis of the entire *gp35* gene. Mutants generated according to the experimental protocol described above are incubated at a high temperature, resulting in the loss of the entire distal half of the tail fiber (*i.e.*, gp35-P36-P37) in the thermolabile mutants. Wild-type phage (and distal half-fibers from thermolabile mutants) are then separated from thermolabile mutant phage that have been inactivated at high temperature (but still have
25 proximal half tail fibers attached) by precipitating both the distal half-fibers and the phage particles containing intact tail fibers with any of the anti-distal half tail-fiber antibodies and protein-A beads. Mutant phage remaining in the supernatant are then reactivated by incubation at low temperature with bacterial extracts containing wild type intact distal half fibers. The thermolabile gene 35 mutants grown at 30°C can be tested for reversible
30 thermolability by inactivation at 60°C and reincubation at 30°C. Inactivation is performed on a concentrated suspension of phage, and reincubation at 30°C is performed either before

or after dilution. If phage are successfully reactivated before, but not after dilution, this indicates that their gp35 is reversibly thermolabile.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those
5 described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

10

15

20

25

30

WHAT IS CLAIMED IS:

1. A composition comprising at least 1 microgram of a purified nondenatured gp35 protein, with the proviso that said composition is not a gel.

5

2. A purified bacteriophage T4 gp35 protein that is not contained in a gel.

3. A purified protein comprising the amino acid sequence depicted in Figure 2 (SEQ ID NO:2) with one or more conservative substitutions relative to said sequence.

10

4. A protein comprising the amino acid sequence depicted in Figure 2 (SEQ ID NO:2) from amino acid residues 1 to 93 with one or more conservative substitutions relative to the sequence in Figure 2.

15

5. A purified protein encoded by a nucleic acid hybridizable to a DNA having a nucleotide sequence consisting of the coding region of SEQ ID NO:1, with the proviso that the protein is not a native gp35 protein.

6. A purified protein comprising an amino acid sequence of 100 amino acids that has at least 60% identity to a gp35 protein having the amino acid sequence depicted in Figure 2 (SEQ ID NO:2).

7. A purified protein comprising at least 8 contiguous amino acids of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 24, and which displays one or more functional activities of a gp35 protein.

8. The protein of claim 7 which is able to be bound by an antibody directed against a gp35 protein.

9. The protein of claim 7 which has only conservative substitutions relative to the sequence in Figure 2 (SEQ ID NO:2).

10. A molecule comprising the protein of claim 7.

11. The protein of claim 6 which specifically binds with the P34 protein oligomer of bacteriophage T4.

5

12. A purified fragment of the protein of claim 4, which comprises at least 8 contiguous amino acids of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 24, and which displays one or more functional activities of a gp35 protein.

10

13. The fragment of claim 12 which is able to be bound by an antibody directed against a gp35 protein.

14. A purified protein variant of a gp35 protein of bacteriophage T4, that is able to
15 be bound by an antibody directed against a gp35 protein, wherein the interaction of said variant with the P36 protein oligomer of bacteriophage T4 is unstable at temperatures between about 40°C and about 60°C.

15. A purified protein variant of a gp35 protein of bacteriophage T4, that is able to
20 be bound by an antibody directed against a gp35 protein, wherein the interaction of said variant with the P34 protein oligomer of bacteriophage T4 is unstable at temperatures between about 40°C and about 60°C.

16. A purified protein variant of a gp35 protein of bacteriophage T4, that (a) is able
25 to be bound by an antibody directed against a gp35 protein, and (b) is conjugated to a group that confers the ability of the variant to bind a ligand.

17. The variant of claim 16, wherein said ligand is selected from the group consisting of avidin, immunoglobulin, and a divalent metal ion.

30

18. A purified molecule comprising a bacteriophage T4 gp35 protein fragment, wherein said fragment consists of at least the amino acid sequence depicted in Figure 2

(SEQ ID NO:2) from amino acids numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79 or 81-93.

19. A purified molecule comprising the amino acid sequence depicted in Figure 2
5 (SEQ ID NO:2) from amino acids numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79 or 81-93, with one or more conservative substitutions relative to said sequence.

20. A purified molecule comprising an amino acid sequence having at least 30%
identity to amino acids numbers 57 to 93 in Figure 2 (SEQ ID NO:2) over a 36 amino acid
10 sequence.

21. A purified protein having at least 60% identity to amino acids numbers 57 to 93
in Figure 2 (SEQ ID NO:2) over a 36 amino acid sequence.

15 22. A purified protein comprising at least a functionally active portion of the amino acid sequence in Figure 2 (SEQ ID NO:2) from amino acids numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-64, 66-79, or 81-93.

23. A purified molecule comprising an amino acid sequence having at least 60%
20 identity to amino acids numbers 1 to 100 in Figure 2 (SEQ ID NO:2) over a 100 amino acid sequence.

24. The purified fragment of claim 7, wherein said fragment lacks at least 10
contiguous amino acids of the sequence depicted in Figure 2 (SEQ ID NO:2).
25

25. A purified nucleic acid, comprising a nucleotide sequence encoding a gp35
protein having the amino acid sequence depicted in Figure 2 (SEQ ID NO: 2), operably
linked to a heterologous promoter that controls expression of the nucleotide sequence.

30 26. A purified nucleic acid, comprising a nucleotide sequence encoding a gp35
protein having the amino acid sequence depicted in Figure 2 (SEQ ID NO: 2), contiguous
with a sequence of at least 10 nucleotides that is not of bacteriophage T4.

27. The purified nucleic acid of claim 25, further comprising nucleotide sequences encoding gp36, gp37 and gp57 proteins, respectively, operably linked to said promoter.

28. The purified nucleic acid of claim 25, in which the nucleic acid is DNA.

5

29. The purified nucleic acid of claim 25, in which the nucleic acid is RNA.

30. A purified nucleic acid comprising a nucleotide sequence absolutely complementary to a nucleotide sequence encoding a gp35 protein having the amino acid sequence depicted in Figure 2 (SEQ ID NO:2), contiguous with a sequence of at least 10 nucleotides that is not of bacteriophage T4.

31. A purified nucleic acid comprising at least 850 contiguous nucleotides of a *gp35* DNA sequence, with the proviso that the nucleic acid does not contain a bacteriophage T4 promoter.

15

32. A purified nucleic acid, comprising a nucleotide sequence encoding a gp35 protein consisting of at least the amino acid sequence shown in Figure 2 from amino acids numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79, or 81-93.

20

33. A purified nucleic acid comprising a nucleotide sequence encoding a protein consisting of at least the amino acid sequence shown in Figure 2 (SEQ ID NO:2) from amino acids numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79 or 81-93, with one or more conservative substitutions relative to said sequence.

25

34. A purified nucleic acid, comprising the nucleotide sequence depicted in Figure 2 (SEQ ID NO:1) from nucleotide numbers 1 to 1,116, wherein said sequence is contiguous to a 3' termination codon.

30

35. A purified nucleic acid, comprising a nucleotide sequence encoding a protein having at least 30% identity to amino acids numbers 57 to 93 in Figure 2 (SEQ ID NO:2) over a 36 amino acid sequence.

36. A purified nucleic acid, comprising a nucleotide sequence encoding a protein containing at least a functionally active portion of the amino acid sequence in Figure 2 from amino acids numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-64, 66-79, or 81-93.

5 37. A purified nucleic acid, comprising a nucleotide sequence encoding the protein of claim 12.

38. The purified nucleic acid of claim 37, wherein said protein is missing at least 10 contiguous amino acids of the sequence depicted in Figure 2 (SEQ ID NO:2).

10

39. A nucleic acid vector comprising the nucleic acid of claim 26 or 33.

40. An expression vector comprising the nucleic acid of claim 33 operably linked to a heterologous promoter that controls expression of the nucleotide sequence in a host cell.

15

41. A host cell that contains the nucleic acid of claim 25.

42. A host cell that contains the nucleic acid of claim 33.

20 43. A host cell that contains the nucleic acid of claim 33 operably linked to a heterologous promoter that controls expression of the nucleotide sequence in the host cell.

44. A method of producing a protein comprising growing the host cell of claim 41 such that the gp35 protein is expressed by the cell, and recovering the expressed protein.

25

45. A method of producing a protein comprising growing the host cell of claim 43 such that the encoded protein is expressed by the cell, and recovering the expressed protein.

46. The product of the method of claim 44.

30

47. The product of the method of claim 45.

48. A kit comprising in one or more containers a pair of nucleic acid primers capable of priming amplification of at least a portion of a gp35 gene, in which the 5' primer is upstream of or comprising a sequence encoding the N-terminus of a gp35 protein.

5

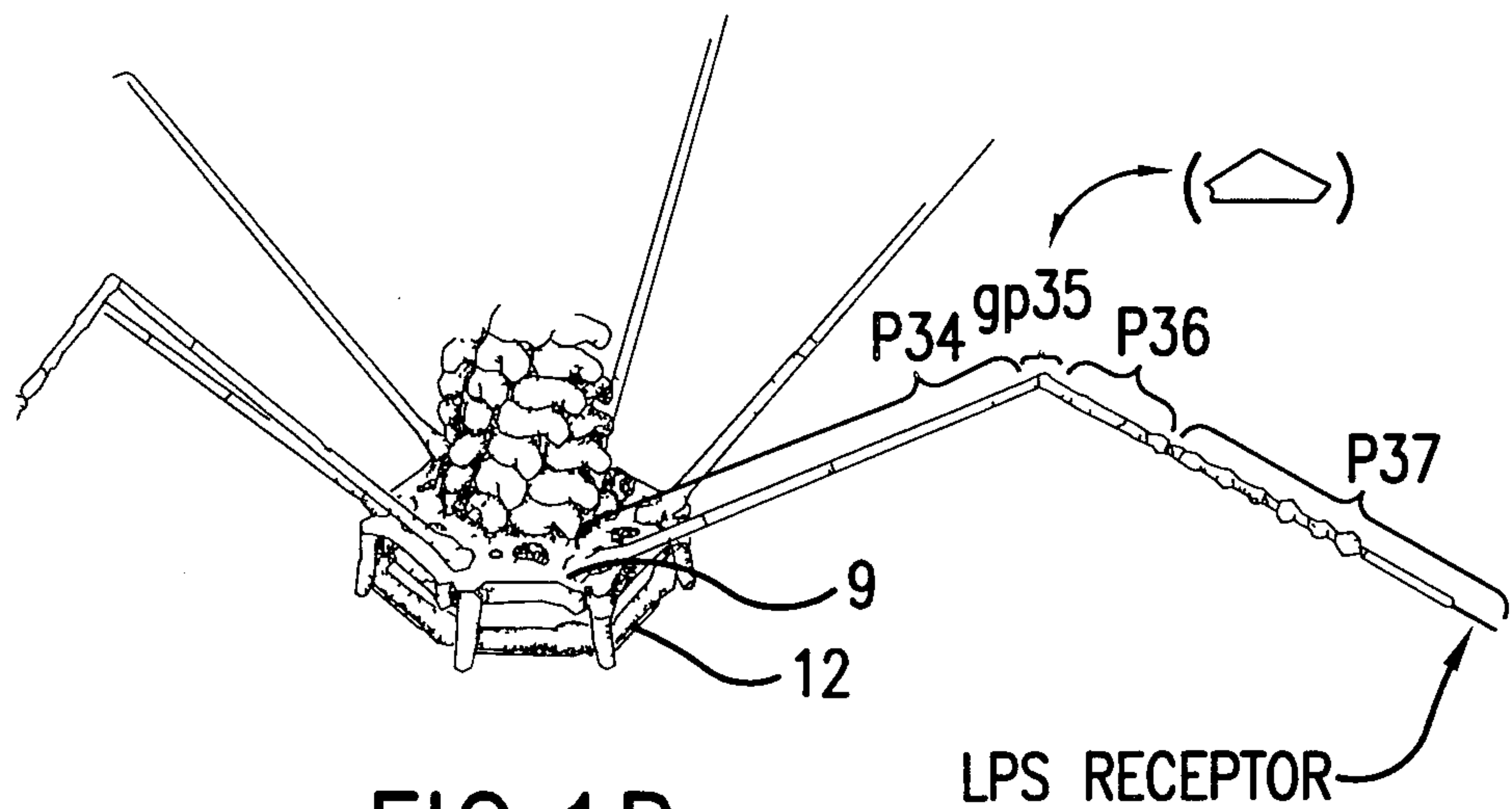
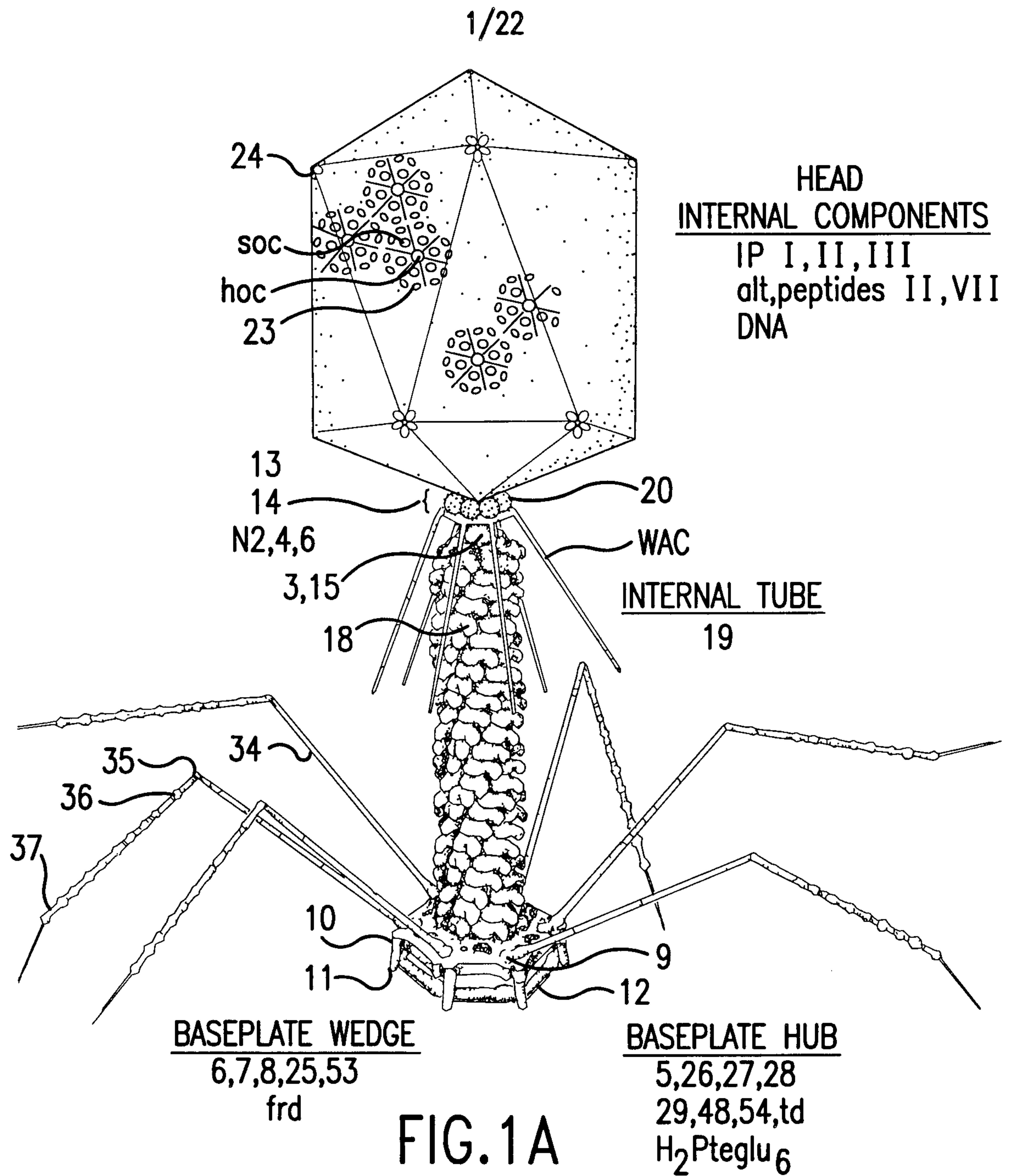
10

15

20

25

30



1 ATG GAA AAA TTT ATG GCC GAG TTT GGA CAA GGA TAT GTC CAA ACG CCA TTT TTA TCG GAA 60
1 M E K F M A E F G C Q G Y V Q T P F L S E 20

61 AGT AAT TCA GTA AGA TAT AAA ATA AGT ATA GCG GGT TCT TCC CCG CTT TCT ACA GCA GGA 120
21 S N S V R Y K I S I A G S C P L S T A G 40

121 CCA TCA TAT GTT AAA TTT CAG GAT AAT CCT GTA GGA AGT CAA ACA TTT ACC GCA GCC CTC 180
41 P S Y V K F Q D N P V G S Q T F S A G L 60

181 CAT TTA AGA GTT TTT GAC CCT TCC ACC GGA CCA TTA GTT GAT AGT AAG TCA TAT GCC TTT 240
61 H L R V F D P S T G A L V D S K S Y A F 80

241 TCC ACT TCA AAT GAT ACT ACA TCA GCT GCT TTT GTT AGT TTC ATG AAT TCT TTG ACG AAT 300
81 S T S N D T T S A A F V S F M N S L T N 100

301 AAT CGA ATT GTT GCT ATA TTA ACT AGT GGA AAG GTT AAT TTT CCT CCT GAA GTA GTA TCT 360
101 N R I V A I L T S G K V N F P P E V V S 120

361 TCG TTA AGA ACC GCC GGA ACG TCT GCC TTT CCA TCT GAT TCT ATA TTG TCA AGA TTT GAC 420
121 W L R T A G T S A F P S D S I L S R F D 140

421 GTA TCA TAT GCT GCT TTT TAT ACT TCT TCT AAA AGA GCT ATC GCA TTA GAG CAT GTT AAA 480
141 V S Y A A F Y T S S K R A I A L E H V K 160

FIG. 2A

481 CTG AGT AAT AGA AAA AGC ACA GAT GAT TAT CAA ACT ATT TTA GAT GTT GTA TTT GAC AGT 540
161 L S N R K S T D D Y Q T I L D V V F D S 180

541 TTA GAA GAT GTA GGG GCT ACC GGG TTT CCA AGA GGA ACG TAT GAA AGT GTT GAG CAA TTC 600
181 L E D V C A T C F P R G T Y E S V E Q F 200

601 ATG TCG GCA GTT GGT GGA ACT AAT GAC GAA ATT GCG AGA TTG CCA ACT TCA GCT GCT ATA 660
201 M S A V C G T N D E I A R L P T S A A I 220

661 AGT AAA TTA TCT GAT TAT AAT TTA ATT CCT GGA GAT GTT CTT TAT CTT AAA GCT CAG TTA 720
221 S K L S D Y N L I P C D V L Y L K A Q L 240

721 TAT GCT GAT GCT GAT TTA CTT GCT CTT GGA ACT ACA AAT ATA TCT ATC CGT TTT TAT AAT 780
241 Y A D A D L L A L C T T N I S I R F Y N 260

781 GCA TCT AAC GGA TAT ATT TCT TCA ACA CAA GCT GAA TTT ACT GGG CAA GCT CGG TCA TCG 840
261 A S N G Y I S S T Q A E F T C Q A C S W 280

841 GAA TTA AAG GAA GAT TAT GTA GTT GTT CCA GAA AAC GCA GTA GGA TTT ACG ATA TAC GCA 900
281 E L K E D Y V V V P E N A V C F T I Y A 300

901 CAG AGA ACT GCA CAA GCT GGC CAA GGT GGC ATG AGA AAT TTA AGC TTT TCT GAA GTA TCA 960
301 Q R T A Q A G Q C G C M R N L S F S E V S 320

FIG. 2B

961	AGA	AAT	GGC	GGC	ATT	TCG	AAA	CCT	GCT	GAA	TTT	GGC	GTC	AAT	GGT	ATT	CGT	GTT	AAT	TAT	1020
321	R	N	G	G	I	S	K	P	A	E	F	G	V	N	G	I	R	V	N	Y	340
1021	ATC	TGC	GAA	TCC	GCT	TCA	CCC	CCG	GAT	ATA	ATG	GTA	CIT	CCT	ACG	CAA	GCA	TCG	TCT	AAA	1080
341	I	C	E	S	A	S	P	P	D	I	M	V	L	P	T	Q	A	S	S	K	360
1081	ACT	GGT	AAA	GTG	TTT	GGG	CAA	GAA	TTT	AGA	GAA	GTT	TAA								1119
361	T	G	K	V	F	C	Q	E	F	R	E	V	*								373

FIG. 2C

1 TAGGAGCCCGGAGA ATG GCC GAG ATT AAA AGA GAA TTC AGA GCA GAA GAT GGT CTG GAC GCA 63
1 M A E I K R E F R A E D G L D A 16

64 GGT GGT GAT AAA ATA ATC AAC GTA CCT TTA CCT GAT CGT ACC GTA GGA ACT GAC GGT GTT 123
17 G G D K I I N V A L A D R T V C T D G V 36

124 AAC GTT GAT TAC TTA ATT CAA GAA AAC ACA GTT CAA CAG TAT GAT CCA ACT CGT GGA TAT 183
37 N V D Y L I Q E N T V Q Q Y D P T R C Y 56

184 TTA AAA GAT TTT GTA ATC ATT TAT GAT AAC CGC TTT TGG GCT GCT ATA AAT GAT AIT CCA 243
57 L K D F V I I Y D N R F W A A I N D I P 76

244 AAA CCA GCA GCA GCT TTT AAT AGC GGA CGC TGG AGA GCA TTA CGT ACC GAT GCT AAC TGG 303
77 K P A G A F N S G R W R A L R T D A N W 96

304 ATT ACG GTT TCA TCT GGT TCA TAT CAA TTA AAA TCT GGT GAA GCA ATT TCG GTT AAC ACC 363
97 I T V S S G S Y Q L K S G E A I S V N T 116

364 GCA GCT GCA AAT GAC ATC ACG TTT ACT TTA CCA TCT TCT CCA ATT GAT GGT GAT ACT AIC 423
117 A A G N D I T F T L P S S P I D C D T I 136

424 GTT CTC CAA GAT ATT GGA GGA AAA CCT GGA GTT AAC CAA GTT TTA ATT GTA GCT CCA GTA 483
137 V L Q D I G G K P G V N Q V L I V A P V 156

FIG. 3A

484	CAA	AGT	ATT	GTA	AAC	TTT	AGA	GGT	GAA	CAG	GTA	CGT	TCA	GTA	CTA	ATG	ACT	CAT	CCA	AAG	543
157	Q	S	I	V	N	F	R	G	E	Q	V	R	S	V	L	M	T	H	P	K	176
544	TCA	CAG	CIA	GTT	TTA	ATT	TTT	AGT	AAT	CGT	CTG	TGG	CAA	ATG	TAT	GTT	GCT	GAT	TAT	AGT	603
177	S	Q	L	V	L	I	F	S	N	R	L	W	Q	M	Y	V	A	D	Y	S	196
604	AGA	GAA	GCT	ATA	GTT	GTA	ACA	CCA	GCG	AAT	ACT	TAT	CAA	GAG	CAA	TCC	AAC	GAT	TTT	ATC	663
197	R	E	A	I	V	V	T	P	A	N	T	Y	Q	A	Q	S	N	D	F	I	216
664	GTA	CGT	AGA	TTT	ACT	TCT	GCT	GCA	CCA	ATT	AAT	GTC	AAA	CIT	CCA	AGA	TTT	GCT	AAT	CAT	723
217	V	R	R	F	T	S	A	A	P	I	N	V	K	L	P	R	F	A	N	H	236
724	GCC	GAT	ATT	ATT	AAT	TTC	GTC	GAT	TTA	GAT	AAA	CTA	AAT	CCG	CIT	TAT	CAT	ACA	ATT	GTT	783
237	G	D	I	I	N	F	V	D	L	D	K	L	N	P	L	Y	H	T	I	V	256
784	ACT	ACA	TAC	GAT	GAA	ACG	ACT	TCA	GTA	CAA	GAA	GTT	GGA	ACT	CAT	TCC	ATT	GAA	GCG	CGT	843
257	T	T	Y	D	E	T	T	S	V	Q	E	V	C	T	H	S	I	E	G	R	296
844	ACA	TCG	ATT	GAC	GGT	TTC	TIG	ATG	TTT	GAT	GAT	AAT	GAG	AAA	TTA	TGG	AGA	CIG	TTT	GAC	903
277	T	S	I	D	G	F	L	M	F	D	D	N	E	K	L	W	R	L	F	D	296
904	GCG	GAT	AGT	AAA	GCG	CGT	TTA	CGT	ATC	ATA	ACG	ACT	AAT	TCA	AAC	ATT	CGT	CCA	AAT	GAA	963
297	G	D	S	K	A	R	L	R	I	I	T	T	N	S	N	I	R	P	N	E	316

FIG. 3B

964	GAA	GTT	ATG	GTA	TTT	GGT	CCC	AAT	AAC	GGA	ACA	ACT	CAA	ACA	ATT	GAG	CTT	AAG	CTT	CCA	1023
317	E	V	M	V	F	G	A	N	G	T	T	Q	T	I	I	E	L	K	L	P	336
1024	ACT	AAT	ATT	TCT	GTT	GGT	GAT	ACT	GTT	AAA	ATT	TCC	ATG	AAT	TAC	ATG	AGA	AAA	GGA	CAA	1083
337	T	N	I	S	V	G	D	T	V	K	I	S	M	N	Y	M	R	K	G	Q	356
1084	ACA	GTT	AAA	ATC	AAA	GCT	GCT	GAT	GAA	GAT	AAA	ATT	GCT	ICT	ICA	GTT	CAA	TIG	CTG	CAA	1143
357	T	V	K	I	K	A	A	D	E	D	K	I	A	S	S	V	Q	L	L	Q	376
1144	TTC	CCA	AAA	CGC	TCA	GAA	TAT	CCA	CCT	GAA	GCT	GAA	TGG	GTT	ACA	GTT	CAA	GAA	TIA	GTT	1203
377	F	P	K	R	S	E	Y	P	P	E	A	E	W	V	T	V	Q	E	L	V	396
1204	TTT	AAC	GAT	GAA	ACT	AAT	TAT	GTT	CCA	GTT	TTG	GAG	CIT	GCT	TAC	ATA	GAA	GAT	ICT	GAT	1263
397	F	N	D	E	T	N	Y	V	P	V	L	E	L	A	Y	I	E	D	S	D	416
1264	GGA	AAA	TAT	IGG	GTT	GTA	CAG	CAA	AAC	GTT	CCA	ACT	GTA	GAA	AGA	GTA	GAT	TCT	TIA	AAT	1323
417	G	K	Y	W	V	V	Q	Q	N	V	P	T	V	E	R	V	D	S	L	N	436
1324	GAT	TCT	ACT	AGA	GCA	AGA	TTA	GGC	GTA	ATT	GCT	TTA	GCT	ACA	CAA	GCT	CAA	GCT	AAT	GTC	1383
437	D	S	T	R	A	R	L	G	V	I	A	L	A	T	Q	A	Q	A	N	V	456
1384	GAT	TTA	GAA	AAT	TCT	CCA	CAA	AAA	GAA	TTA	GCA	ATT	ACT	CCA	GAA	ACG	TTA	GCT	AAT	CGT	1443
457	D	L	E	N	S	P	Q	K	E	L	A	I	T	P	E	T	L	A	N	R	476

FIG. 3C

8/22

1444 ACT GCT ACA GAA ACT CGC AGA GGT ATT GCA AGA ATA GCA ACT ACT GCT CAA GTG AAT CAG 1503
477 T A T E T R R C I A R I A T T A Q V N Q 496

1504 AAC ACC ACA TTC TCT TTT GCT GAT GAT ATT ATC ATC ACT CCT AAA AAG CTG AAT GAA AGA 1563
497 N T T F S F A D D I I I I T P K K L N E R 516

1564 ACT GCT ACA GAA ACT CGT AGA GGT GTC GCA GAA ATT GCT ACG CAG CAA GAA ACT AAT GCA 1623
517 T A T E T R R C V A E I A T Q Q E T N A 536

1624 GGA ACC GAT GAT ACT ACA ATC ATC ACT CCT AAA AAG CTT CAA GCT CGT CAA GGT TCT GAA 1683
537 G T D T T I I I T P K K L Q A R Q G S E 556

1744 GAA TTA AAT GGT ACG AAT GTT TAT AAT AAA AAC ACT GAT AAT TTA GTT GTT TCA CCT AAA 1803
577 E L N G T N V Y N K N T D N L V V S P K 596

1804 GCT TTG GAT CAG TAT AAA GCT ACT CCA ACA CAG CAA GGT CCA GTA ATT TTA GCA GTT CAA 1863
597 A L D Q Y K A T P T Q Q C A V I L A V E 616

1864 AGT GAA GTA ATT GCT GGA CAA AGT CAG CAA GCA TCG GCA AAT GCT GTT GTA ACG CCA GAA 1923
617 S E V I A C Q S Q Q C W A N A V V T P E 636

1924 ACG TTA CAT AAA AAG ACA TCA ACT GAT GGA AGA ATT GGT TTA ATT GAA ATT GCT ACG CAA 1983
637 T L H K K T S T D C R I C L I E I A T Q 656

FIG. 3D

1984 AGT GAA GTT AAT ACA GGA ACT GAT TAT ACT CGT GCA GTC ACT CCT AAA ACT TTA AAT GAC 2043
657 S E V N T G T D Y T R A V T P K T L N D 676

2044 CGT AGA GCA ACT GAA AGT TTA AGT CGT ATA GCT GAA ATT GCT ACA CAA GTT GAA TTC GAC 2103
677 R A T E S L S G I A E I A T Q V E F D 696

2104 GCA GGC GTC GAC GAT ACT CGT ATC TCT ACA CCA TTA AAA ATT AAA ACC AGA TTT AAT AGT 2163
697 A C V D T R I S T P L K I K T R F N S 716

2164 ACT GAT CGT ACT TCT GTT GCT CTA TCT GGA TTA GTT GAA TCA GGA ACT CTC TGG GAC 2223
717 T D R T S V V A L S G L V E S G T L W D 736

2224 CAT TAT ACA CTT AAT ATT CTT GAA GCA AAT GAG ACA CAA CGT GGT ACA CTT CGT GTA GCT 2283
737 H Y T L N I L E A N E T Q R G T L R V A 756

2284 ACG CAG GTC GAA GCT GCT GCG GGA ACA TTA GAT AAT GTT TTA ATA ACT CCT AAA AAG CTT 2343
757 T Q V E A A C T L D N V L I T P K K L 776

2344 TTA GGT ACT AAA TCT ACT GAA GCG CAA GAG GGT GTT ATT AAA GTT GCA ACT CAG TCT GAA 2403
777 L C T K S T E A Q E C V I K V A T Q S E 796

2404 ACT GTG ACT GGA ACG TCA GCA AAT ACT GCT GTA TCT CCA AAA AAT TTA AAA TGG ATT GCG 2463
797 T V T C T S A N T A V S P K N L K W I A 816

FIG. 3E

2464 CAG AGT GAA CCT ACT TGG GCA GCT ACT ACT GCA ATA AGA GGT TTT GTT AAA ACT TCA TCT 2523
817 Q S E P T W A A T T A I R G F V K T S S 836

2524 GGT TCA ATT ACA TTC GTT GGT AAT GAT ACA GTC GGT TCT ACC CAA GAT TTA GAA CTG TAT 2583
837 G S I T F V G N D T V G S T Q D L E L Y 856

2584 GAG AAA AAT AGC TAT GCG GTA TCA CCA TAT GAA TTA AAC CGT GTA TTA GCA AAT TAT TTG 2643
857 E K N S Y A V S P Y E L N R V L A N Y L 876

2644 CCA CTA AAA GCA AAA GCT GCT GAT ACA AAT TTA TTG GAT GGT CIA GAT TCA TCT CAG TTC 2703
877 P L K A K A A D T N L L D G L D S S Q F 896

2704 ATT CGT AGG GAT ATT GCA CAG ACG GTT AAT GGT TCA CIA ACC TTA ACC CAA CAA ACG ATT 2763
897 I R R D I A A Q T V N G S L T L T Q Q T N 916

2764 CTG AGT GCC CCT CTT GTA TCA ICT AGT ACT GGT GAA TTT GGT GGT TCA TTG GCC GCT AAT 2823
917 L S A P L V S S S T G E F G C S L A A N 936

2824 AGA ACA TTT ACC ATC CGT AAT ACA GGA GCC CCG ACT AGT ATC GTT TTC GAA AAA GGT CCT 2883
937 R T F T I R N T G A P T S I V F E K G P 956

2884 GCA TCC GCG GCA AAT CCT GCA CAG TCA ATG AGT ATT CGT GTA TGG GGT AAC CAA TTT GGC 2943
957 A S G A N P A Q S M S I R V W G N Q F G 976

FIG. 3F

2944 GCC GGT AGT GAT ACC ACC CGT ICC ACA GTG TTT GAA GTT GGC GAT GAC ACA TCT CAT CAC 3003
977 G G S D T T R S T V F E V G D D T S H H 996

3004 TTT TAT TCT CAA CGT AAT AAA GAC GGT AAT ATA GCG TTT AAC ATT AAT GGT ACT GTA ATG 3063
997 F Y S Q R N K D G N I A F N I N G T V M 1016

3064 CCA ATA AAC ATT AAT CCT ICC GGT TTG ATG AAT GTG AAT GGC ACT GCA ACA TTC GGT CGT 3123
1017 P I N I N A S G L M N V N G T A T F G R 1036

3124 TCA GTT ACA GCC AAT GGT GAA TTC ATC AGC AAG TCT GCA AAT GCT TTT AGA GCA ATA AAC 3183
1037 S V T A N C E F I S K S A N A F R A I N 1056

3184 GGT GAT TAC GGA TTC TTT ATT CGT AAT GAT GCC TCT AAT ACC TAT TTT TTG CTC ACT GCA 3243
1057 G D Y C F I R N D A S N T Y F L L T A 1076

3244 GCC GGT GAT CAG ACT GGT GGT TTT AAT GGA TTA CGC CCA TTA ATT AAT AAT CAA TCC 3303
1077 A C D Q T C G C F N G L R P L L I N N Q S 1096

3304 GGT CAG ATT ACA ATT GGT GAA GGC TTA ATC ATT GCC AAA GGT GTT ACT ATA AAT TCA GGC 3363
1097 C Q I T I C E C L I I A K G V T I N S G 1116

3364 GGT TTA ACT GTT AAC TCG AGA ATT CGT TCT CAG GGT ACT AAA ACA TCT GAT TTA TAT ACC 3423
1117 G L T V N S R I R S Q C T K T S D L Y T 1136

FIG. 3G

12/22

3424	CGT	GGC	CCA	ACA	TCT	GAT	ACT	GTA	GGA	TTC	TGG	TCA	ATC	GAT	ATT	AAT	GAT	TCA	GCC	ACT	3483
1137	R	A	P	T	S	D	T	V	G	F	W	S	I	D	I	N	D	S	A	T	1156
3484	TAT	AAC	CAG	TTC	CCG	GGT	TAT	TTT	AAA	ATG	GTT	GAA	AAA	ACT	AAT	GAA	GTG	ACT	GGG	CTT	3543
1157	Y	N	Q	F	P	G	Y	F	K	M	V	E	K	T	N	E	V	T	G	L	1176
3544	CCA	TAC	TTA	GAA	CGT	GGC	GAA	GAA	GTT	AAA	TCT	CCT	GGT	ACA	CTG	ACT	CAG	TTT	GGT	AAC	3603
1177	P	Y	L	E	R	G	E	V	K	S	P	G	T	L	T	Q	F	G	N		1196
3604	ACA	CTT	GAT	TCG	CTT	TAC	CAA	GAT	TGG	ATT	ACT	TAT	CCA	ACG	ACG	CCA	GAA	GCG	CGT	ACC	3663
1197	T	L	D	S	L	Y	Q	D	W	I	T	Y	P	T	T	P	E	A	R	T	1216
3664	ACT	CGC	TGG	ACA	CGT	ACA	TGG	CAG	AAA	ACC	AAA	AAC	TCT	TGG	TCA	AGT	TTT	GTT	CAG	GTA	3723
1217	T	R	W	T	R	T	W	Q	K	T	K	N	S	W	S	S	F	V	Q	V	1236
3724	TTT	GAC	GGA	GGT	AAC	CCT	CCT	CAA	CCA	TCT	GAT	ATC	GGT	GCT	TTA	CCA	TCT	GAT	AAT	GCT	3783
1237	F	D	G	C	N	P	P	Q	P	S	D	I	G	A	L	P	S	D	N	A	1256
3784	ACA	ATG	GGG	AAT	CTT	ACT	ATT	CGT	GAT	TTC	TIG	CGA	ATT	GGT	AAT	GTT	CGC	ATT	GTT	CCT	3843
1257	T	M	G	N	L	T	I	R	D	F	L	R	I	G	N	V	R	I	V	P	1276
3844	GAC	CCA	GTG	AAT	AAA	ACG	GTT	AAA	TTT	GAA	TGG	GTT	GAA	TAA	GAGGTATT	ATG	GAA	AAA	TTT		3905
1277	D	P	V	N	K	T	V	K	F	E	W	V	E	*		M	E	K	F	4	

FIG. 3H

3906 ATG GCC GAG ATT TGG ACA AGG ATA TGT CCA AAC GCC ATT TTA TCG GAA AGT AAT TCA GTA 3965
5 M A E I W T R I C P N A I L S E S N S V 24

3966 ACA TAT AAA ATA AGT ATA GCG GGT TCT TGC CCG CTT TCT ACA GCA GGA CCA TCA TAT GTT 4025
25 R Y K I S I A G S C P L S T A G P S Y V 44

4026 AAA TTT CAG GAT AAT CCT GTA GGA AGT CAA ACA TTT AGG CGC AGG CCT TCA TTT AAG AGT 4085
45 K F Q D N P V G S Q T F R R P S F K S 64

4086 TTT TGA CCCTTCCACCGGACCATTAGTGTAGTAACTCAT ATG CTT TTT CGA CTT CAA ATG ATA CTA 4153
65 F * M L F R L Q M I L 9

4154 CAT CAG CTG CTT TIG TTA GTT TTC ATG AAT TCT TTG ACC AAT AAT CGA ATT GTT GCT ATA 4213
10 H Q L L L L V F M N S L T N N R I V A I 29

4214 TTA ACT AGT GGA AAG GTT AAT TTT CCT CCT GAA GTA GTA TCT TCG TTA AGA ACC GCC GGA 4273
30 L T S G K V N F P P E V V S W L R T A G 49

4274 ACG TCT GCC TTT CCA TCT GAT TCT ATA TTG TCA AGA TTT GAC GTA TCA TAT GCT GCT TTT 4333
50 T S A F P S D S I L S R F D V S Y A A F 69

4334 TAT ACT TCT AAA AGA GCT ATC GCA TTA GAG CAT GTT AAA CTG AGT AAT AGA AAA AGC 4393
70 Y T S S K R A I A L E H V K L S N R K S 89

FIG. 3I

4394 ACA GAT GAT TAT CAA ACT ATT TTA GAT GTT GTA TTT GAC AGT TTA GAA GAT GTA GGA GCT 4453
90 T D D Y Q T I L D V V F D S L E D V G A 109

4454 ACC GCG TTT CCA AGA AGA ACG TAT GAA AGT GTT GAG CAA TTC ATG TCG GCA GTT GCT GGA 4513
110 T G F P R R T Y E S V E Q F M S A V G G 129

4514 ACT AAT AAC GAA ATT GCG AGA TTG CCA ACT TCA GCT GCT ATA AGT AAA TTA TCT GAT TAT 4573
130 T N N E I A R L P T S A A I S K L S D Y 149

4574 AAT TTA ATT CCT GGA GAT GTT CTT TAT CTT AAA GCT CAG TTA TAT GCT GAT GCT GAT TTA 4633
150 N L I P G D V L Y L K A Q L Y A D A D L 169

4634 CTT GCT CTT GGA ACT ACA AAT ATA TCT ATC CGT TTT TAT AAT GCA TCT AAC GGA TAT ATT 4693
170 L A L G T T N I S I R F Y N A S N G Y I 189

4694 TCT TCA ACA CAA GCT GAA TTT ACT GCG CAA GCT GCG TCA TGG GAA TTA AAG GAA GAT TAT 4753
190 S S T Q A E F T G Q A C S W E L K E D Y 209

4754 GTA GTT GTT CCA GAA AAC GCA GTA GGA TTT ACG ATA TAC GCA CAG AGA ACT GCA CAA GCT 4813
210 V V V P E N A V G F T I Y A Q R T A Q A 229

4814 GGC CAA GGT GGC ATG AGA AAT TTA AGC TTT TCT GAA GTA TCA AGA AAT GGC GGC ATT TCG 4873
230 G Q G G M R N L S F S E V S R N C G I S 249

FIG. 3J

4874 AAA CCT GCT GAA TTT GGC GTC AAT GGT ATT CGT GTT AAT TAT AIC TGC GAA TCC GCT TCA 4933
250 K P A E F G V N G I R V N Y I C E S A S 269

4934 CCT CCG GAT ATA ATG GTA CTT CCT ACG CAA GCA TCG TCT AAA ACT GTG AAA GTG TTT GCG 4993
270 P P D I M V L P T Q A S S K T G K V F C 289

4994 CAA GAA TTT AGA GAA GTT TAA ATTGAGGACCCCTTCGGTTCCTTTTCTTTATAAATACIATTCAAAATAAA 5066
290 Q E F R E V * 296

5067 GCGGCAIACA ATG GCT GAT TTA AAA GTA GGT TCA ACA ACT CGA GGC TCT GTC ATT TCG CAT 5127
1 M A D L K V G S T T G C S V I W H 17

5128 CAA GCA AAT TTT CCA TTG AAT CCA GCC GGT GAC GAT GTA CTC TAT AAA TCA TTT AAA ATA 5187
18 Q G N F P L N P A G D D V L Y K S F K I 37

5188 TAT TCA GAA TAT AAC AAA CCA CCA GCT GCT GAT AAC GAT TTC GTT TCT AAA GCT AAT GCT 5247
38 Y S E Y N K P Q A A D N D F V S K A N C 57

5248 GGT ACT TAT GCA TCA AAG GTA ACA TTT AAC GCT GGC ATT CAA GTC CCA TAT GCT CCA AAC 5307
58 G T Y A S K V T F N A G I Q V P Y A P N 77

5308 ATC ATG ACC CCA TCC GCG ATT TAT GCG GGT AAC GGT GAT GGT GCT ACT TTT GAT AAA GCA 5367
78 I M S P C G I Y G G N G D G A T F D K A 97

FIG. 3K

5368 AAT ATC GAT ATT GTT TCA TGG TAT GGC GTA GGA TTT AAA TCG TCA TTT GTT TCA ACA GGC 5427
98 N I D I V S W Y G V G F K S S F C S T C 117

5428 CGA ACT GTT GTA ATT AAT ACA CGC AAT GGT GAT ATT AAC ACA AAA GGT GTT GTG TCG GCA 5487
118 R T V V I N T R N G D I N T K G V V S A 137

5488 GCT GGT CAA GTA AGA AGT GGT GCG GCT CCT ATA GCA GCG AAT GAC CTT ACT AGA AAG 5547
138 A G Q V R S G A A A P I A A N D L T R K 157

5548 GAC TAT GTT GAT GGA GCA ATA AAT ACT GTT ACT GCA AAT GCA AAC TCT AGG GTG CTA CCG 5607
158 D Y V D G A I N T V T A N A N S R V L R 177

5608 TCT GGT GAC ACC ATG ACA GGT AAT TTA ACA GCG CCA AAC TTT TTC TCG CAG AAT CCT GCA 5667
178 S G D T M T G N L T A P N F F S Q N P A 197

5668 TCT CAA CCC TCA CAC GTT CCA CGA TTT GAC CAA ATC GTA ATT AAG GAT TCT GTT CAA GAT 5727
198 S Q P S H V P R F D Q I V I K D S V Q D 217

5728 TTC GGC TAT TAT TAA GAGGACIT ATG GCT ACT TTA AAA CAA ATA CAA TTT AAA AGA AGC AAA 5789
218 F G Y Y * M A T L K Q I Q F K R S K 13

5790 ATC GCA GGA ACA CGT CCT GCT TCA GTA TTA GCC GAA GGT GAA TTG GCT ATA AAC TTA 5849
14 I A C T R P A A S V L A E G E L A I N L 33

FIG. 3L

5850 AAA GAT AGA ACA ATT TTT ACT AAA GAT GAT TCA GGA AAT ATC GAT CTA GGT TTT GAT 5909
34 K D R T I F T K D D S G N I I D L C F A 53

5910 AAA GGC GCG CAA GTT GAT GGC AAC GTT ACT ATT AAC GGA CTT TTG AGA TTA AAT GGC GAT 5969
54 K C G Q V D G N V T I N G L L R L N C D 73

5970 TAT GTA CAA ACA GGT GGA ATG ACT GTA AAC GGA CCC ATT GGT TCT ACT GAT GGC GTC ACT 6029
74 Y V Q T C G M T V N G P I G S T D C V T 93

6030 GGA AAA ATT TTC AGA TCT ACA CAG GGT TCA TTT TAT GCA AGA GCA ACA AAC GAT ACT TCA 6089
94 G K I F R S T Q C S F Y A R A T N D T S 113

6090 AAT GCC CAT TTA TGG TTT GAA AAT GCC GAT GGC ACT GAA CGT GGC GTT ATA TAT GCT CGC 6149
114 N A H L W F E N A D C G T E R C V I Y A R 133

6150 CCT CAA ACT ACA ACT GAC GGT GAA ATA CGC CTT AGG GTT AGA CAA GGA ACA GGA AGC ACT 6209
134 P Q T T T D G E I R L R V R Q C T C S T 153

6210 GCC AAC AGT GAA TTC TAT TTC CGC TCT ATA AAT GGA GGC GAA TTT CAG GCT AAC CGT ATT 6269
154 A N S E F Y F R S I N C G E F Q A N R I 173

6270 TTA GCA TCA GAT TCG TTA GTA ACA AAA CGC ATT GCG GTT GAT ACC GTT ATT CAT GAT GCC 6329
174 L A S D S L V T K R I A V D T V I H D A 193

FIG. 3M

6330 AAA GCA TTT GGA CAA TAT GAT TCT CAC TCT TTG GTT AAT TAT GTT TAT CCT GCA ACC GGT 6389
194 K A F C Q Y D S H S L V N Y V Y P G T G 213

6390 GAA ACA AAT GGT GTA AAC TAT CTT CGT AAA GTT CGC GCT AAG TCC GGT GGT ACA ATT TAT 6449
214 E T N C V N Y L R K V R A K S G G T I Y 233

6450 CAT GAA ATT GTT ACT GCA CAA ACA GGC CTG GCT GAT GAA GTT TCT TCG TCG TCT GGT GAT 6509
234 H E I V T A Q T C L A D E V S W S G D 253

6510 ACA CCA GTA TTT AAA CTA TAC GGT ATT CGT GAC GAT GGC AGA ATG ATT ATC CGT AAT AGC 6569
254 T P V F K L Y C I R D D G R M I I R N S 273

6570 CTT GCA TTA GGT ACA TTC ACT ACA AAT TTC CCG TCT AGT GAT TAT GGC AAC GTC GGT GTA 6629
274 L A L C T F T T N F P S S D Y G N V G V 293

6630 ATG GGC GAT AAG TAT CTT GTT CIC GGC GAC ACT GTA ACT GGC TTG TCA TAC AAA AAA ACT 6689
294 M G D K Y L V L G D T V T G L S Y K K T 313

6690 GGT GTA TTT GAT CTA GTT GGC GGT GGA TAT TCT GTT GCT TCT ATT ACT CCT GAC AGT TTC 6749
314 G V F D L V C G C Y S V A S I T P D S F 333

6750 CGT AGT ACT CGT AAA GGT ATA TTT GGT CGT TCT GAG GAC CAA GGC GCA ACT TGG ATA ATG 6809
334 R S T R K C I F C R S E D Q C A T W I M 353

FIG. 3N

6810 CCT GGT ACA AAT GCT GCT CTC TTG ICT GTT CAA ACA GCT GAT AAT AAC AAT GCT GGA 6869
354 P G T N A A L L S V Q T Q A D N N N A C 373

6870 GAC GGA CAA ACC CAT ATC GGG TAC AAT GCT GGG GGT AAA ATG AAC CAC TAT TTC CGT GGT 6929
374 D C Q T H I C Y N A C G K M N H Y F R C 393

6930 ACA GGT CAG ATG AAT ATC AAT ACC CAA CAA GGT ATG GAA ATT AAC CCG GGT ATT TTG AAA 6989
394 T C Q M N I N T Q Q C M E I N P C I L K 413

6990 TTG GTA ACT GGC TCT AAT AAT GTA CAA TTT TAC GCT GAC GGA ACT ATT TCT TCC ATT CAA 7049
414 L V T C S N N V Q F Y A D C T I S S I Q 433

7050 CCT ATT AAA TTA GAT AAC GAG ATA TTT TTA ACT AAA TCT AAT AAT ACT GCG GGT CTT AAA 7109
434 P I K L D N E I F L T K S N N T A C G L K 453

7110 TTT GGA GCT CCT AGC CAA GTT GAT GGC ACA AGG ACT ATC CAA TGG AAC GGT GGT ACT CGC 7169
454 F G A P S Q V D G T R T I Q W N C G T R 473

7170 GAA GGA CAG AAT AAA AAC TAT GTG ATT ATT AAA GCA TGG GGT AAC TCA TTT AAT GCC ACT 7229
474 E G Q N K N Y V I I K A W C N S F N A T 493

7230 GGT GAT AGA TCT CGC GAA ACG GTT TTC CAA GTA TCA GAT AGT CAA GGA TAT TAT TTT TAT 7289
494 G D R S R E T V F Q V S D S Q C Y Y F Y 513

FIG. 30

7290 GCT CAT CGT AAA GCT CCA ACC GGC GAC GAA ACT ATT GGA CGT ATT GAA GCT CAA TTT GCT 7349
514 A H R K A P T C D E T I C R I E A Q F A 533

7350 GCG GAT GTT TAT GCT AAA GGT ATT ATT GCC AAC GGA AAT TTT AGA GTT GTT GCG TCA AGC 7409
534 G D V Y A K G I I A N G N F R V V G S S 553

7410 GCT TTA GCC GGC AAT GTT ACT ATG TCT AAC GGT TTG TTT GTC CAA GGT GGT TCT TCT ATT 7469
554 A L A G N V T M S N G L F V Q C G S S I 573

7470 ACT GGA CAA GTT AAA ATT GGC GGA ACA GCA AAC CTG AGA ATT TGG AAC GCT GAA TAT 7529
574 T G Q V K I G C T A N A L R I W N A E Y 593

7530 GGT GCT ATT TTC CGT CGT TCG GAA AGT AAC TTT TAT ATT ATT CCA ACC AAT CAA AAT GAA 7589
594 G A I F R R S E S N F Y I I P T N Q N E 613

7590 GGA GAA AGT GGA GAC ATT CAC AGC TCT TTG AGA CCT GTG AGA ATA GGA TTA AAC GAT GGC 7649
614 G E S G D I H S S L R P V R I C L N D C 633

7650 ATG GTT GCG TTA GCA AGA GAT TCT TTT ATA GTA GAT CAA AAT AAT GCT TTA ACT ACG ATA 7709
634 M V C L G R D S F I V D Q N N A L T T I 653

7710 AAC AGT AAC TCT CCG ATT AAT GCC AAC TTT AGA ATG CAA TTG GCG CAG TCG GCA TAC ATT 7769
654 N S N S R I N A N F R M Q L C Q S A Y I 673

FIG. 3P

7770 GAT GCA GAA TGT ACT GAT GCT GTT CCC GCG GGT GCA GGT TCA TTT GCT TCC CAG AAT 7829
674 D A E C T D A V R P A G A C S F A S Q N 693

7830 AAT GAA GAC GTC CGT CCG TTC TAT ATG AAT ATT GAT AGA ACT GAT GCT AGT GCA TAT 7889
694 N E D V R A P F Y M N I D R T D A S A Y 713

7890 GTT CCT ATT TTG AAA CAA CGT TAT GTT CAA GGC AAT GGC TGC TAT TCA TTA GGG ACT TTA 7949
714 V P I L K Q R Y V Q G N G C Y S L G T L 733

7950 ATT AAT AAT GGT AAT TTC CGA GTT CAT TAC CAT GGC GGC GGA GAT AAC GGT TCT ACA GGT 8009
734 I N N G N F R V H Y H G C G D N G S T G 753

8010 CCA CAG ACT GCT GAT TTT GGA TCG GAA TTT ATT AAA AAC GGT GAT TTT ATT TCA CCT CGC 8069
754 P Q T A D F G W E F I K N G D F I S P R 773

8070 GAT TTA ATA GCA GGC AAA GTC AGA TTT GAT AGA ACT GGT AAT ATC ACT GGT GGT TCT GGT 8129
774 D L I A G K V R F D R T G N I T G C S G 793

8130 AAT TTT GCT AAC TTA AAC AGT ACA ATT GAA TCA CTT AAA ACT GAT ATC AIG TCG AGT TAC 8189
794 N F A N L N S T I E S L K T D I M S S Y 813

8190 CCA ATT GCT GCT CCG ATT CCT TCG CCG AGT GAT TCA GTT CCT GCT GGA TTT GCT TTG AIG 8249
814 P I G A P I P W P S D S V P A C F A L M 833

FIG. 3Q

8250 GAA GGT CAG ACC TTT GAT AAG ICC GCA TAT CCA AAG TTA GCT GTT GCA TAT CCT ACC GGT 8309
834 E G Q T F D K S A Y P K L A V A Y P S G 853

8310 GTT ATT CCA GAT ATG CGC GCG CAA ACT ATC AAG GGT AAA CCA AGT GGT GCT GCT TTG 8369
854 V I P D M R G Q T I K G K P S G R A V L 873

8370 AGC GCT GAG GCA GAT GGT GTT AAG GCT CAT AGC CAT AGT GCA TCG GCT TCA AGT ACT GAC 8429
874 S A E A D G V K A H S H S A S A S T D 893

8430 TTA GGT ACT AAA ACC ACA TCA ACC TTT GAC TAT GGT ACC AAG GGA ACT AAC AGT ACC GGT 8489
894 L G T K T T S S F D Y G T K G T N S T G 913

8490 GGA CAC ACT CAC TCT GGT AGT GGT TCT ACT AGC ACA AAT GGT GAG CAC AGC CAC TAC ATC 8549
914 G H T H S G S G S T S T N G E H S H Y I 933

8550 GAG GCA TGG AAT GGT ACT GGT GTA GGT GGT AAT AAG ATG TCA TCA TAT GCC ATA TCA TAC 8609
934 E A W N G T G V G C N K M S S Y A I S Y 953

8610 AGG GCG GGT AGT AAC ACT AAT GCA GCA GCG AAC CAC AGT CAC ACT TTC TCT TTT GGG 8669
954 R A G G S N T N A A C N H S H T F S F G 973

8670 ACT AGC AGT GCT GCG GAC CAT ICC CAC TCT GTA GGT ATT GGT GCT CAT ACC CAC ACC GTA 8729
974 T S S A G D H S H S V G I G A H T H T V 993

8730 GCA ATT GGA TCA CAT GGT CAT ACT ATC ACT GTA AAT AGT ACA GGT AAT ACA GAA AAC ACG 8789
994 A I G S H G H T I T V N S T C N T E N T 1013

8790 GTT AAA AAC ATT GCT TTT AAC TAT ATC GTT CGT TTA GCA TAA GGAGAGGGCTTCGGCCCTTCTAA 8855
1014 V K N I A F N Y I V R L A * 1027

FIG. 3R