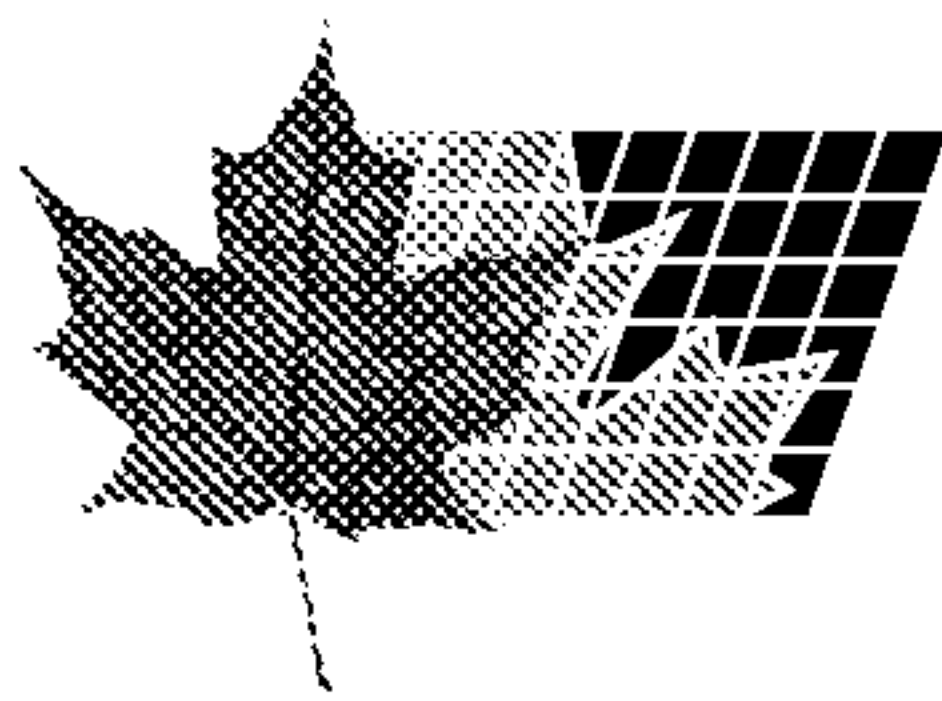




(72) MASIGNANI, VEGA, IT

(72) RAPPUOLI, RINO, IT

(72) PIZZA, MARIAGRAZIA, IT

(72) SCARLATO, VINCENZO, IT

(72) GRANDI, GUIDO, IT

(71) CHIRON S.P.A., IT

(51) Int.Cl.⁶ C12N 15/31, C07K 14/22, C07K 16/12, A61K 39/095

(30) 1997/11/06 (9723516.2) GB

(30) 1997/11/14 (9724190.5) GB

(30) 1997/11/18 (9724386.9) GB

(30) 1997/11/27 (9725158.1) GB

(30) 1997/12/10 (9726147.3) GB

(30) 1998/01/14 (9800759.4) GB

(30) 1998/09/01 (9819016.8) GB

(54) **ANTIGENES DE NEISSERIA**

(54) **NEISSERIAL ANTIGENS**

(57) Cette invention concerne des protéines provenant de *Neisseria meningitidis* (souches A & B) et de *Neisseria gonorrhoeae*, y compris des séquences d'acides aminés, les séquences nucléotidiques correspondantes, des données d'expression et des données sérologiques. Ces protéines sont des antigènes utiles pour des vaccins, des compositions immunogéniques, et/ou des tests de diagnostic.

(57) The invention provides proteins from *Neisseria meningitidis* (strains A and B) and from *Neisseria gonorrhoeae* including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)	A2	(11) International Publication Number: WO 99/24578 (43) International Publication Date: 20 May 1999 (20.05.99)																					
(21) International Application Number: PCT/IB98/01665 (22) International Filing Date: 9 October 1998 (09.10.98) (30) Priority Data: <table border="0" style="width: 100%;"> <tr> <td style="width: 30%;">9723516.2</td> <td style="width: 40%;">6 November 1997 (06.11.97)</td> <td style="width: 30%;">GB</td> </tr> <tr> <td>9724190.5</td> <td>14 November 1997 (14.11.97)</td> <td>GB</td> </tr> <tr> <td>9724386.9</td> <td>18 November 1997 (18.11.97)</td> <td>GB</td> </tr> <tr> <td>9725158.1</td> <td>27 November 1997 (27.11.97)</td> <td>GB</td> </tr> <tr> <td>9726147.3</td> <td>10 December 1997 (10.12.97)</td> <td>GB</td> </tr> <tr> <td>9800759.4</td> <td>14 January 1998 (14.01.98)</td> <td>GB</td> </tr> <tr> <td>9819016.8</td> <td>1 September 1998 (01.09.98)</td> <td>GB</td> </tr> </table> (71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT). (72) Inventors; and (75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbucco, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).		9723516.2	6 November 1997 (06.11.97)	GB	9724190.5	14 November 1997 (14.11.97)	GB	9724386.9	18 November 1997 (18.11.97)	GB	9725158.1	27 November 1997 (27.11.97)	GB	9726147.3	10 December 1997 (10.12.97)	GB	9800759.4	14 January 1998 (14.01.98)	GB	9819016.8	1 September 1998 (01.09.98)	GB	(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB). (81) Designated States: 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, 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, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, 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 <i>Without international search report and to be republished upon receipt of that report.</i>
9723516.2	6 November 1997 (06.11.97)	GB																					
9724190.5	14 November 1997 (14.11.97)	GB																					
9724386.9	18 November 1997 (18.11.97)	GB																					
9725158.1	27 November 1997 (27.11.97)	GB																					
9726147.3	10 December 1997 (10.12.97)	GB																					
9800759.4	14 January 1998 (14.01.98)	GB																					
9819016.8	1 September 1998 (01.09.98)	GB																					
(54) Title: NEISSERIAL ANTIGENS (57) Abstract The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.																							

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET
COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2

NOTE: Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE
THAN ONE VOLUME

THIS IS VOLUME OF

NOTE: For additional volumes-please contact the Canadian Patent Office

WO 99/24578

PCT/IB98/01665

-1-

NEISSERIAL ANTIGENS

This invention relates to antigens from *Neisseria* bacteria.

BACKGROUND ART

Neisseria meningitidis and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

WO 99/24578

PCT/IB98/01665

-2-

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

WO 99/24578

PCT/IB98/01665

-3-

porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

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

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide
10 sequences disclosed in the examples.

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

Nucleic acid comprising fragments of these sequences are also provided. These should comprise
15 at least n consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, n is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

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

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

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

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

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

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

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the
10 presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

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

According to further aspects, the invention provides various processes.

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

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

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

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

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

5 General

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

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

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

Definitions

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

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

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

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

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

Expression systems

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

i. Mammalian Systems

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

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

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

in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

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

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

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

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal

viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

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

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

20 ii. Baculovirus Systems

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

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

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

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

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

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

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

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

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

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

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter

WO 99/24578

PCT/IB98/01665

-13-

and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion
5 can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

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

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

30

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

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

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

iii. Plant Systems

20 There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in
25 Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by
30 gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*,. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

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

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

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

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

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

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

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

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

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

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

iv. Bacterial Systems

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

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

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac*
30 promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21].

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA
5 polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for
10 the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979)
15 "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

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

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the
25 N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene
30 [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.*

(1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

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

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

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

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

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

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

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

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

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

Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; Streptococcus lividans [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], Streptomyces lividans [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, Campylobacter], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, Staphylococcus], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, Streptococcus].

v. Yeast Expression

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

of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

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

10 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters
15 which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA*
20 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

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

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

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

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

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

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

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

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

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

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

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

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See eg. [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*];

[Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent
 5 Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 Saccharomyces]; [Beach and Nurse (1981) *Nature* 300:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; Yarrowia].

Antibodies

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

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

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

recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

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

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

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

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

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

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

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

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

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

Delivery Methods

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

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

Vaccines

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

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

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

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

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

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

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

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

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

20 Gene Delivery Vehicles

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

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can
30 also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,

picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

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

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

15 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

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

25 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; 5 WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 10 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors 15 employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. 20 Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in 25 which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted 30 terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

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

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

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

- 5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

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

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

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA*

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

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

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

15 Delivery Methods

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

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

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

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

Polynucleotide and polypeptide pharmaceutical compositions

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

A. Polypeptides

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

B. Hormones, Vitamins, etc.

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

C. Polyalkylenes, Polysaccharides, etc.

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

D. Lipids, and Liposomes

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

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary

but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta*. 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

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

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

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

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta*
 25 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

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

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

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

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

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

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

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

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

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

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

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

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

25 Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

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

Nucleic Acid Hybridisation

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

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

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes
5 can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

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

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

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

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

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for
30 a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology,

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

Nucleic Acid Probe Assays

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

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

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some
20 variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe
25 sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as
30 temperature, salt condition and the like. For example, for diagnostic applications, depending on the

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

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

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

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

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

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

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (◆) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psорт.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

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

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

10 In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8).
15 After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one ChCl₃/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70%
20 ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A
25 sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included

a *XhoI* restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *BamHI-XhoI* or *EcoRI-XhoI*), and pET21b+ (using either *NdeI-XhoI* or *NheI-XhoI*).

5'-end primer tail: CGCGGATCCCATATG (*BamHI-NdeI*)

5 CGCGGATCCGCTAGC (*BamHI-NheI*)

CCGGAATTCTAGCTAGC (*EcoRI-NheI*)

3'-end primer tail: CCCGCTCGAG (*XhoI*)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *XhoI* primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (*NdeI*)

5'-end primer tail: CGGGATCC (*BamHI*)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *NheI-BamHI* restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (*NheI*)

3'-end primer tail: CGGGATCC (*BamHI*)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH₄OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD₂₆₀ was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl₂), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10µl DMSO or 50µl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

WO 99/24578

PCT/IB98/01665

-52-

	95°C	65-70°C	72°C
--	------	---------	------

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

- 5 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.
- 10 **D) Digestion of PCR fragments**

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- 15 – *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *Sall/PstI* for cloning into pGex-His and further expression of
- 20 the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or

25 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the
10 vector pTRC99 (Pharmacia).

F) Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer.
15 The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boehringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then
20 centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin
25 Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SalI* or, for ORF 122, *SalI-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth
5 with 50µl/ml ampicillin.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described
10 above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for
15 induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by
20 OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were
25 diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again.
30 The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia)

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD₂₈₀ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

- 5 For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation
 10 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM
 15 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D₂₈₀ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D₂₈₀ was 0.1. 21µl of each
 20 fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-
 25 14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

10 M) Mice immunisations

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)₃ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)₃, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

20 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA was considered positive when OD₄₉₀ was 2.5 times the respective pre-immune sera.

10 **O) FACScan bacteria Binding Assay procedure.**

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

15 R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₆₂₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₆₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and purification results.

Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
20  201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
25  451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501 AGACCG...
```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAQGN AAAQYNLGAM
51  YXQRTVRVRD DAEAVRWYRQ PAEQGLAQAO YNLGWMYANG RXVRQDDTEA
30  101 VRWYRQAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
151 AQNNLGVMYA ERXRVRQD...
```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
35  101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
40  351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501 AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
551 ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA
```

WO 99/24578

PCT/IB98/01665

-61-

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```

1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
51 YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAA YNLGWMYANG RGVRRDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
5  151 AQNNLGVMYA ERGRVQRDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY*
```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
10  151 TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201 GCTTGGCAAG GCTTGTCAA ACGGATACCA AGACAGCTGC GACAATGACC
251 AACGCCTGAA AGCGGGTTAT TGA
```

This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```

1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
15  51 YAERRGVRRD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *
```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

```

20  orf37.pep      10      20      30      40      50      60
      MKQTVXMLAAALIALGLNRPVWXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVRVRD
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf37a      10      20      30      40      50      60
      MKQTVKWLAAALIALGLNQA VVWADDVSDFR ENLQAAAQGNAAAQNNLGVMYAERRGVRRD

25  orf37.pep      70      80      90      100     110     120
      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
      | | : | : : |
      orf37a      70      80      90
      RALAQEWLGKACQNGYQDSCDNDQRLKAGYX
```

30 Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATTCAATTT GGGCGTGATG
151 TATGAAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
35  201 GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTTGG
251 GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TCGCTCGCT
301 CAACAATGGC TTGGCAAGGC TTGTCAAAC GGAGACCAA ACAGCTGCGA
351 CAATGACCAA CGCCTGAAGG CGGGTTATTA A
```

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```

40  1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
      51 YENGQGVRRD YVQAVQWYRK ASEQGDAAQ YNLGLMYDYG RGVRRDLALA
      101 QQWLKGACQN GDQNSCDNDQ RLKAGY*
```

The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

```

45  orf37.pep      MKQTVXMLAAALIALGLNRPVWXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVRVRD  60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf37ng      MKQTVKWLAAALIALGLNQA VVWAGDVSDFR ENLQAAEQGNAAAQFNLGVMYENGQGVRRD  60

50  orf37.pep      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG  120
      : ||| : ||| : ||| ||||| || : || |||| : | : | : |
      orf37ng      YVQAVQWYRKASEQGDAAQAYNLGLMYDGRGVRRDLALAAQWLKGACQNGDQNSCDNDQ  120

      orf37.pep      VIYAEGRGVRQDDVEAVRWFRQAAAQGVAAQNNLGVMYAERXVRVRD  168
55  orf37ng      RLKAGY 126
```

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

		10	20	30	40	50	60
	orf37-1.pep	MKQTVKWLAAALIALGLNRAVWADDVSD	FRENLQAAAQGNAAQYNLGAMYYKGRGVRRD				
5	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSD	FRENLQAAEQGNAAQFNLGVMYENGQGVRRD				
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf37-1.pep	DAEAVRWYRQAAEQGLAQYNLGWMYANGRGVRRD	DDTEAVRWYRQAAAQGVVQAQYNLG				
	orf37ng	YVQAVQWYRKASEQGDAAQYNLGLMYDGRGVRRD	-----				
		70	80	90			
15	orf37-1.pep	VIYAEGRGVRRDDVEAVRWFRQAAAQGVAAQNNLGVMYAERRGVRRD	DRALAQEWFGKAC				
	orf37ng	-----	-----LALAQWLKAC				
			100				
20		190	199				
	orf37-1.pep	QNGDQDGCNDNDQRLKAGYX					
	orf37ng	QNGDQNSCNDNDQRLKAGYX					
		110	120				

Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

	TTCGGCGA	CATCGGCGGT	TTGAAGGTCA	ATGCCCCCGT	CAAATCCGCA
40	GGCGTATTGG	TCGGGCGCGT	CGGCGCTATC	GGACTTGACC	CGAAATCCTA
	TCAGGCGAGG	GTGCGCCTCG	ATTTGGACGG	CAAGTATCAG	TTCAGCAGCG
	ACGTTTCCGC	GCAAATCCTG	ACTTCsGGAC	TTTTGGGCGA	GCAGTACATC
	GGGCTGCAGC	AGGGCGGCGA	CACGGA AAAC	CTTGCTGCCG	GCGACACCAT
	CTCCGTAACC	AGTTCTGCAA	TGGTTCTGGA	AAACCTTATC	GGCAAATTCA
45	TGACGAGTTT	TGCCGAGAAA	AATGCCGACG	GCGGCAATGC	GGAAAAAGCC
	GCCGAATAA				

This corresponds to the amino acid sequence <SEQ ID 10>:

1	FGDIGGLKVN	APVKSAGVLV	GRVGAIGLDP	KSYQARVRLD	LDGKYQFSSD
51	VSAQILTSGL	LGEQYIGLQQ	GGDTENLAAG	DTISVTSSAM	VLENLIGKFM

WO 99/24578

PCT/IB98/01665

-63-

101 TSFAEKNADG GNAEKAAE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

```

5      20      30      40      50      60      70
ybrd.h LGIGALVFLGLRVANVQGFAETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
      10      20      30

10     80     90     100     110     120     130
ybrd.h KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
N.m      KSYQARVRDLDDGKY-QFSSDVSAQILTSGLLGEQYIGLQQG---GDTENLAAGDTISVT
      40     50     60     70     80

15     140     150     160
ybrd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
20     N.m      SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX
      90     100     110     120

```

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

```

25     20     30     40     50     60     70
ybrd      GAAVAFLAFRVAGGAFFGGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
      10     20     30

30     80     90     100     110     120     130
ybrd      KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
N.m      KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
      40     50     60     70     80     90

35     140     150     160
ybrd      VLENLIGKFMTSFAEKNADGGNAEKAAEX
N.m      VLENLIGKFMTSFAEKNADGGNAEKAAEX
      100     110     120

```

The complete ybrd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```

50      1      ..ATTTTGATAT ACCTCATCCG CAAGAATCTA GGTTGCCCCG TCTTCTTCTT
      51      TCAGGAACGC CCCGGAAGG ACGGAAAACC TTTTAAAATG GTCAAATTCC
      101     GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
      151     GAACGCCTGA CACCGTTCGG CAAAAAACTG CGTGCCGcCA GTwTGACGA
      201     ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
      251     CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAAC
      301     CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG

```

WO 99/24578

PCT/IB98/01665

-64-

5 351 GCGCAACGCG CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
 401 TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
 451 AAAAAAGTAT TAATCAAGGA AGGGATTTCC GCACAGGGCG AACAA.aCCAT
 501 GCCCCCTTTC ACAGGAAAAC GCAAACCTCGC CGTCGTCCGT GCGGGCGGAC
 551 ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA
 601 ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
 651 CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
 701 ACGTCGCCGT CGCCGTCCGC AACAAACCGCA TCCGCCGCCA AATCGCCGAA
 751 AAAGCCGCCG CGCTCGGCTT CGCCCTGCCG GTACTGGTTC ATCCGGACGC
 10 801 GACCGTCTCG CCTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
 851 AAGCGGTCG..

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

15 1 ..ILIIYLIRKNL GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG
 51 ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN
 101 RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTV
 151 KKVLIKEGIS AQGEXTMPFF TGRKRLAVVG AGGHGKVVAD LAAALGRYRE
 201 IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
 251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

20 1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
 51 ACTGATTTTC CTCTCGCCAG TATTTTTGAT TTTGATATAC CTCATCCGCA
 101 AGAATCTAGG TTCGCCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
 151 GGAAAACCTT TTAAATGGT CAAATTCCGT TCCATGCGCG ACGCGCTTGA
 201 TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
 25 251 AAAAAGTGGC TGCCGCCAGT TTGGACGAAC TGCCTGAATT ATGGAATATC
 301 TTAAAGGCG AGATGAGCCT GGTCGGCCCC CGCCCGCTGC TGATGCAATA
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
 451 GAAAAATTTC CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT
 30 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
 551 GGATTTCGCG ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
 601 AAACTCGCCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT
 651 TGCCGCCGCA CTCGGCCGGT ACAGGGAAAT CGTTTTTCTG GACGACCGCG
 701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
 35 751 GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
 851 CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
 901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCTGAC AGGCAGGCAG
 951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
 40 1001 ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCG
 1051 GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
 1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
 1201 AAGCCGCTGC CGCGCAAAAA CCCCAGAGACC TCGACAGCAT AA

45 This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

50 1 MSKFFKRLFD IVASASGLIF LSPVFLILIIY LIRKNLGSPV FFFQERPGKD
 51 GKPFKMVKFR SMRDALDS DG IPLPDGERLT PFGKKLRAAS LDELPELWNI
 101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD
 151 EKFAADVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPFFTGKR
 201 KLAVVGAGGH GKVVADLAAA LGRYREIVFL DDRAQGSVNG FSVIGTLLLL
 251 ENSLSPEQYD VAVAVGNNRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT
 301 VGQGSVVMK AVVQAGSVLK DGVIIVNTAAT VDHDCLLNAF VHISPGAHLS
 351 GNTHIGEESE IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
 401 KPLPRKNPET STA*

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:

10

20

30

WO 99/24578

PCT/IB98/01665

-65-

	orf3.pep	ILIIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR
	orf3a	MSKFFKRLFDIVASASGLIFLSPVFLILIIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR
5		10 20 30 40 50 60
	orf3.pep	SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL
	orf3a	SMHDAALDSGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL
10		40 50 60 70 80 90
	orf3.pep	YDNFQNRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCLDIKILLTVKKVL
15	orf3a	YDNFQNRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSLCLDIKILLTVKKVL
		130 140 150 160 170 180
	orf3.pep	IKEGISAQGEATMPFFTGKRKLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRQGSVNG
20	orf3a	IKEGISAQGEATMPFFTGKRKLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRQGSVNG
		160 170 180 190 200 210
	orf3.pep	FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT
25	orf3a	FPVIGTTLLLENSLSPEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLVHPDSTVSPSAT
		220 230 240 250 260 270
30	orf3.pep	VGQGSVVMKAV
	orf3a	VGQGGVMAKAVVQADSVLKDGIVNTAATVDHDCLLDAFVHISPGAHLNTRIGEEESW
		280 310 320 330 340 350 360

35 The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

	1	ATGAGTAAAT	TCTTCAAACG	CCTGTTTGAC	ATTGTTGCCT	CCGCCTCGGG
	51	ACTGATTTTC	CTCTCGCCAG	TATTTTGTAT	TTTGATATAC	CTCATCCGCA
	101	AGAATCTGGG	TTCGCCCGTC	TTCTTCTTTC	AGGAACGCCC	CGGAAAGGAC
40	151	GGAAAACCTT	TTAAAATGGT	CAAATTCGGT	TCCATGCACG	ACGCGCTTGA
	201	TTCAGACGGC	ATTCTGCTGC	CCGACGGAGA	ACGCCTGACA	CCGTTCGGCA
	251	AAAAACTGCG	TGCCGCCAGT	TTGGACGAAC	TGCCCCGAAC	GTGGAACGTC
	301	CTCAAAGGCG	ACATGAGCCT	GGTCGGCCCC	CGCCCCGCTG	TGATGCAATA
	351	TCTGCCGCTG	TACGACAAC	TCCAAAACCG	CCGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
45	451	GAACGCTTCG	CATGCGACAT	CTGGTATATC	GACCACTTCA	GCCTGTGCCT
	501	CGACATCAAA	ATCCTACTGC	TGACGGTTAA	AAAAGTATTA	ATCAAAGAAG
	551	GGATTTCCGC	ACAGGGCGAA	GCCACCATGC	CCCCTTTCAC	AGGAAAACGC
	601	AAACTTGCCG	TCGTCGGTGC	GGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCG
50	701	TCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
	751	GAAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCGCCGTCG	CCGTGCGCAA
	801	CAACCGCATC	CGCCGCCAAA	TCGCCGAAAA	AGCCGCCGCG	CTCGGCTTCG
	851	CCCTGCCCGT	CCTGATTCAT	CCGGACTCGA	CCGTCTCGCC	TTCTGCAACA
	901	GTCGGACAAG	GCGGCGTCGT	TATGGCGAAA	GCCGTCGTAC	AGGCTGACAG
55	951	CGTATTGAAA	GACGGCGTAA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
	1001	ATTGCCTGCT	TGATGCTTTC	GTCCACATCA	GCCCCGGCGC	GCACCTGTCG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCTGG	ATAGGCACAG	GCGCGTGCAG
	1101	CCGCCAGCAG	ATCCGTATCG	GCAGCCGCGC	AACCATTGGA	GCGGGCGCAG
	1151	TCGTCGTGCG	CGACGTTTCA	GACGGCATGA	CCGTGCGGGG	CAACCCGGCA
60	1201	AAACCATTGG	CAGGCAAAAA	TACCGAGACC	CTGCGGTCGT	AA

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

	1	MSKFFKRLFD	IVASASGLIF	LSPVFLILII	LIRKNLGSPV	FFFQERPGKD
	51	GKPFKMVKFR	SMHDAALDSG	ILLPDGERLT	PFGKKLRAAS	LDELPELWNV
65	101	LKGDMSLVGP	RPLLMQYLPL	YDNFQNRHE	MKPGITGWAQ	VNGRNALSWD
	151	ERFACDIWYI	DHFSLCLDIK	ILLTVKKVL	IKEGISAQGE	ATMPFFTGKR
	201	KLAVVGAGGH	GKVVAELAAA	LGTYGEIVFL	DDRQGSVNG	FPVIGTTLLL
	251	ENSLSPEQFD	IAVAVGNNRI	RRQIAEKAAA	LGFALPVLH	PDSTVSPSAT

WO 99/24578

PCT/IB98/01665

-66-

301 VGQGGVMAK AVVQADSVLK DGIVVNTAAT VDHDCLLDAF VHISPGAHLS
 351 GNTRIGEESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
 401 KPLAGKNTET LRS*

Two transmembrane domains are underlined.

5 ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

		10	20	30	40	50	60
	orf3a.pep	MSKFFKRLFDIVASASGLIFLSPVFLILIYLRKNLGSPPVFFFQERPGKDGKPFKMKVKFR					
10	orf3-1	MSKFFKRLFDIVASASGLIFLSPVFLILIYLRKNLGSPPVFFFQERPGKDGKPFKMKVKFR					
		10	20	30	40	50	60
	orf3a.pep	70	80	90	100	110	120
		SMHDALDSGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL					
		:					
15	orf3-1	SMRDALDSGILPLDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL					
		70	80	90	100	110	120
	orf3a.pep	130	140	150	160	170	180
		YDNFQNRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSCLLDIKILLTVKKVL					
20	orf3-1	YDNFQNRHEMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLLDIKILLTVKKVL					
		130	140	150	160	170	180
	orf3a.pep	190	200	210	220	230	240
		IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRVOGSVNG					
25	orf3-1	IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKVVAELAAALGRYREIVFLDDRAQGSVNG					
		190	200	210	220	230	240
	orf3a.pep	250	260	270	280	290	300
		FPVIGTTLLLENSLSPEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLHPDSTVSPSAT					
30	orf3-1	FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
		250	260	270	280	290	300
	orf3a.pep	310	320	330	340	350	360
		VGQGGVMAKAVVQADSVLKDGIVVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEESW					
		:					
40	orf3-1	VGQGSVMAKAVVQAGSVLKDGIVVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEESW					
		310	320	330	340	350	360
	orf3a.pep	370	380	390	400	410	
		IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLAGKNTETLRSX					
45	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
		370	380	390	400	410	

Homology with hypothetical protein encoded by yvfc gene (accession Z71928) of *B. subtilis*

ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

50	ORF3	3	IYLIRKNLGSPPVFFFQERPGKDGKPFKMKVKFRSMRDGLYSDGIPLDGERLTPFGKKLRA	62
			I ++R +GSPVFF Q RPG GKPF + KFR+M D S G LPD RLT G+ +R	
	yvfc	27	IHAVRLKIGSPVFFKQVRPGLHGKPFITYKFRMTDERDSKGNLLPDEVRLTKTGRLIRK	86
	ORF3	63	ASXDELPELWNILKGEMSLVGPRPLLMQYLPLYDNFQNRHEMKPGITGWAQVNGRNALS	122
55			S DELP+L N+LKG++SLVGPRPLLM YLPLY Q RRHE+KPGITGWAQ+NGRNA+S	
	yvfc	87	LSIDELPQLLNVLKGDLSLVGPRPLLM DYLPYTEKQARRHEVKPGITGWAQINGRNAIS	146
	ORF3	123	WDEKFCADVWYIDHFSCLLDXXXXXXXXXXXXXXXXXEGISAQGEEXTMPFFTG	172
			W++KF DVWY+D++S LD EGI T FTG	
60	yvfc	147	WEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEGIQQTNHVTAERFTG	196

WO 99/24578

PCT/IB98/01665

-67-

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N. gonorrhoeae*:

5	orf3	ILIIYLIRKNLGSPVFFFOERPGKDGKPFKMVKFR	34
	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKNKGSPVFFIRERPGKDGKPFKMVKFR	60
10	orf3	SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL	94
	orf3ng	SMRDALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL	120
15	orf3	YDNFQNRHHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCCLDIKILLTVKKVL	154
	orf3ng	YNKFQNRHHEMKPGITGWAQVNGRNALSWDEKFSFCDVWYTDNFSFWLDMKILFLTVKKVL	180
20	orf3	IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng	IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAEALAAALGTGEIVFLDDRTQGSVNG	240
25	orf3	FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	274
	orf3ng	FPVIGTTLLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFKLPVLIHPDATVSPSAI	300
25	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLNTRIGEEER	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

30	1	ATGAGTAAAG	CCGTCAAACG	CCTGTTTCGAC	ATCATCGCAT	CCGCATCGGG
	51	GCTGATTGTC	CTGTGCGCCG	TGTTTTTGGT	TTTAATATAC	CTCATCCGCA
	101	AAACTTAGG	TTCGCCCGTC	TTCTTCattc	GCGAACGCCc	cgGAAAGGAc
	151	ggaaaacCTT	TTAAAATGGT	CAAATTCCGT	TCCAtgcgcg	acgcgcttGA
	201	TTCAGACGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTCGGCA
35	251	AAAAATTACG	CGCCACCACT	TTGGACGAAC	TTCTGAATT	ATGGAATGTC
	301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
	351	TCTGCCGCTT	TACAACAAAT	TTCAAACCG	CCGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
	451	GAAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTCA	GCTTTTGGCT
40	501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
	551	GCATTTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTTCG	GGGGAATCGC
	601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
	701	CCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
45	751	GAAAACAGTT	TATCGCCCGA	ACAATTTCGAC	ATCACCGTCG	CCGTCGGCAA
	801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
	851	AACTGCCCGT	TCTGATTCAT	CCCAGCGCGA	CCGTCTCGCC	TTCTGCAATA
	901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCTGAC	AGGCCGGCAG
	951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
50	1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCGGGCGC	GCACCTGTCG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
	1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgccgGT	GCAGGGgcGG
	1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTCGCGGG	CAACCCGGCA
	1201	AAGCCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence <SEQ ID 18>:

55	1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKNLGSPV	FFIRERPGKD
	51	GKPFKMVKFR	SMRDALDSDG	IPLPDSERLT	DFGKKLRATS	LDELPELWNV
	101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRHE	MKPGITGWAQ	VNGRNALSWD
	151	EKFSCDVWYT	DNFSFWLDMK	ILFLTVKKVL	IKEGISAQGE	ATMPPFAGNR
	201	KLAVIGAGGH	GKVVAEALAA	LGTYGEIVFL	DDRTQGSVNG	FPVIGTTLLL
60	251	ENSLSPEQFD	ITVAVGNNRI	RRQITENAAA	LGFKLPVLIH	PDATVSPSAI
	301	IGQGSVVMKAV	AVVQAGSVLK	DGVIVNTAAT	VDHDCLLDAF	VHISPGAHLN
	351	GNTRIGEEER	IGTGACSRQQ	TTVGSVGTAG	AGAVIVCDIP	DGMTVAGNPA
	401	KPLTGKNPKT	GTA*			

WO 99/24578

PCT/IB98/01665

-68-

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

		10	20	30	40	50	60
	orf3-1.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNL	GSPVFFFQERPGKD	GKPFKMKVKFR			
5	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLRKKNL	GSPVFFIRERPGKD	GKPFKMKVKFR			
		10	20	30	40	50	60
	orf3-1.pep	SMRDALDSDGIPLPDGERLTPFGKKLRAASLDEL	PELWNILKGEMSLV	GPRPLLMQYLPL			
10	orf3ng	SMRDALDSDGIPLPDSERLTDFGKKLRATSLDEL	PELWNVLKGEMSLV	GPRPLLMQYLPL			
		70	80	90	100	110	120
	orf3-1.pep	YDNFQNRHMKPGITGWAQVNGRNALSWDEKFACD	VWYIDHFSCLDIK	ILLTVKKVL			
15	orf3ng	YNKFQNRHMKPGITGWAQVNGRNALSWDEKFS	CDVWYTDNFSFWL	DMKILFLTVKKVL			
		130	140	150	160	170	180
	orf3-1.pep	IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKV	VADLAAALGRYRE	IVFLDDRAQGSVNG			
20	orf3ng	IKEGISAQGEATMPFFAGNRKRLAVIGAGGHGKV	VAELAAALGTYGE	IVFLDDRTQGSVNG			
		190	200	210	220	230	240
	orf3-1.pep	FSVIGTTLLENSLSPEQYDVA	VAVGNNRIRRQIAE	KAAALGFALPVL	VHPDATVSPSAT		
25	orf3ng	FPVIGTTLLENSLSPEQFDIT	VAVGNNRIRRQITEN	AAALGFKLPVL	IHPDATVSPSAI		
		250	260	270	280	290	300
	orf3-1.pep	VGQGSVMAKAVVQAGSVL	KDGVIVNTAATVDH	DCLLNAFVHIS	PGAHL	SGNTHIGEE	SW
35	orf3ng	IGQGSVMAKAVVQAGSVL	KDGVIVNTAATVDH	DCLLDAFVHIS	PGAHL	SGNTRIGEE	SR
		310	320	330	340	350	360
	orf3-1.pep	IGTGACSRQQIRIGSR	ATIGAGAVVVRD	VSDGMTVAGN	PAKPLPRKN	PETSTAX	
40	orf3ng	IGTGACSRQQTTVG	SGVTAGAGAVIV	CDIPDGMTVAGN	PAKPLTGKN	PKTGTAX	
		370	380	390	400	410	

In addition, ORF3ng shows significant homology with a hypothetical protein from *B.subtilis*:

45	gnl PID e238668 (Z71928) hypothetical protein [Bacillus subtilis]
	>gi 1945702 gnl PID e313004 (Z94043) hypothetical protein [Bacillus subtilis]
	>gi 2635938 gnl PID e1186113 (Z99121) similar to capsular polysaccharide biosynthesis [Bacillus subtilis]Length = 202
	Score = 235 bits (594), Expect = 3e-61
	Identities = 114/195 (58%), Positives = 142/195 (72%)
50	Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLRKKNL
	+KRLFD+ A+ L S + L I ++R +GSPVFF + RPG GKPF + KFR+M D
	Sbjct: 3 LKRLFDLTAAIFLLCCTSVIILFTI
55	Query: 65 ALDSDGIPLPDSERLTDFGKKLRATSLDEL
	DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRPLLM YLPLY +
	Sbjct: 63 ERDSKGNLLPDEVRLTKTGRLIRKLSIDELPQLLNVLKGDLSLVGPRPLLM DYLPY
60	Query: 125 QNRHMKPGITGWAQVNGRNALSWDEKF
	Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG
	Sbjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDL
	Query: 185 ISAQGEATMPFFAGN 199
	I T F G+
65	Sbjct: 183 IQQTNHVTAERFTGS 197

The hypothetical product of *yvf*c gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

```

1  ..AACCATATGG CGATTGTCAT CGACGAATAC GCGGCGACAT CCGGCTTGGT
51  CACCTTTGAA GACATCATCG AGCAAATCGT CCGCGAAATC GAAGACGAGT
101 TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301 TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA
15 401 CAGTTTAG

```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51  WRIHAATEIE DINTFFGTEY SIEEADTI XR PGHSRVGTSA RARRKSPYRR
101 FAVHRRTRRQ PPPAYADGDP REVS....XR RFCTV*

```

20 Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```

1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCCG
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTTCCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
25 201 CAGCCGTATG AACGTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAGA GTTCCGCGAA
30 451 CAGCGCAACC ATATGGCGAT TGTCATCGAC GAATACGGCG GCACATCCGG
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601 GAACGCTGGC GCATCCATGC AGTACCGAA ATCGAAGACA TCAACACCTT
651 CTTCCGGCAG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC
35 701 ATTCAAGAGT TGGGACATCT GCCCGTGCGC GGCGAAAAAG TCCTTATCGG
751 CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC
801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
851 TGACGGTACG GGCGTTTTCT GTTTCAATCC GCCCCATCCG CCAAACATAA

```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```

40 1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLRLLE
51  KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVF FEDIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEEADTI RPHSRVGTSA ARARRKSPYR
45 251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCCG
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
50 151 AAAGTCCTCG ATTTTCTCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201 CAGCCGTATG AACGTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGTGAAGAC
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCCGTCT

```


orf5ng-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVDADTLTRLEKVLDFAELEV

	orf5-1	 MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 10 20 30 40 50 60
5	orf5ng-1.pep	70 80 90 100 110 120 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN
	orf5-1	RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN 70 80 90 100 110 120
10	orf5ng-1.pep	130 140 150 160 170 180 EQFHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG :
	orf5-1	EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG 130 140 150 160 170 180
15	orf5ng-1.pep	190 200 210 220 230 240 DIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSGIGT : : : : :
20	orf5-1	EIEDEFDEDDSDNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT 190 200 210 220 230
25	orf5ng-1.pep	250 260 270 280 290 300 PARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACTAVSAQFRMTVRSFSVSIRP : :
	orf5-1	SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVS---TAVSAQFRMTVRAFSVSIRP 240 250 260 270 280 290
30	orf5ng-1.pep	IRQTX
	orf5-1	IRQTX 300

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

Homology with hemolysin homolog TlyC (accession U32716) of *H.influenzae*

ORF5 and TlyC proteins show 58% aa identity in 77 aa overlap (BLASTp).

40

ORF5	2	HMAIVIDEYGGTSGLVTFEDIIEQIVGEIEDEFDEDDSDADNIHAVSSDTWRIHAATEIED	61
		HMAIV+DE+G SGLVT EDI+EQIVG+IEDEFDE++ AD I +S T+ + A T+I+D	
TlyC	166	HMAIVVDEFGAVSGLVTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD	224
ORF5	62	INTFFGTEYSIEEADTI	78
		N F T++ EE DTI	
TlyC	225	FNAQFNTDFDDEEVDTI	241

45 ORF5ng-1 also shows significant homology with TlyC:

SCORES Initl: 301 Initn: 419 Opt: 668
Smith-Waterman score: 668; 45.9% identity in 242 aa overlap

```

50      orf5ng-1.pep      MDGAQPKTNFFERLIARLAR-EPDSAEDVLNLLRQAHEQEVEFDADTLTRLEK
                        | ||: |:|: : | : |:~::~||:~::~::| :| :|
tlyc_haein MNDEQQNSNQSENTKKPFFQSLSFGRFFQGELKNREELVEVIRDSEQNDLIDONTREMIEG
                        10         20         30         40         50         60

55      orf5ng-1.pep      VLDFAELEVVDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGE--DKDEVLGILH
                        60         70         80         90        100        109
tlyc_haein VMEIAELVRDIMIPRSQIIFIEDQQDLNTCLNTIIESAHSRFPVIADADDRDNIVGILH
                        70         80         90        100        110        120

60      orf5ng-1.pep      AKDLLKYMF-NPEQHFLKSVLRAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGL
                        110        120        130        140        150        160
tlyc_haein AKDLLKFLREDAEVFDLSSLLRPVVIVPESKRVRMLKDFRSERFHMAIVVDEF GAVSGL

```

WO 99/24578

PCT/IB98/01665

-73-

```

                    130      140      150      160      170      180
                    170      180      190      200      210      220
5  orf5ng-1.pep VTFEDIIEQIVGDIIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEAD
   tlyc_haein  VTIEDILEQIVGDIIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQFNTDFDDEEVD
                    190      200      210      220      230
10 orf5ng-1.pep TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACTAVSAQF
   tlyc_haein  TIGGLIMQTFGYLPKRGEIILKNLQFKVTSADSRRLIQLRVTPDEHLAEMNNVDEKSE
                    240      250      260      270      280      290

```

15 Homology with a hypothetical secreted protein from *E.coli*:

ORF5a shows homology to a hypothetical secreted protein from *E.coli*:

```

20 sp|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION
   >gi|1778577 (U82598) similar to H. influenzae [Escherichia coli] >gi|1786879
   (AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
   approx. 440 aa protein YTFL_HAEIN SW: P44717 [Escherichia coli] Length = 292

```

Score = 212 bits (533), Expect = 3e-54
 Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

```

25 Query: 2   DGAQPKTNFXRLIARLAR-EPDSAEDVLTLLRQAHEQEVDADTLLRLEKVLDFSDLEV 60
   D   K   F   L+++L   EP + +++L L+R + + ++ D DT   LE V+D +D   V
   Sbjct: 10  DTISNKKGFFSLLLSQLFHGEPKNRDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

30 Query: 61  RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
   RD MI RS+M   LK N +++   +I++AHSRFPVI EDKD + GIL AKDLL +M   +
   Sbjct: 70  RDIMIPRSQMITLKRNQTLDDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

35 Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTTFEDIIEQIV 179
   E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
   Sbjct: 130 AEAFSMDKVLVRQAVVPESKRVDRLKEFRSQRYHMAIVIDEFGGVSGGLVTIEDILELIV 189

   Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADT 229
   G+IEDE+DE++ D   +S   W + A   IED N   FGT +S EE DT
   Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVDT 238

```

40 Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 5

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

```

1  ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTT
51 GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC

```

WO 99/24578

PCT/IB98/01665

-74-

5
10
101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
151 GATGCCTTCA GCGGCAATCC TGAAGGCGAG TTTTTCCTCCG ACAGCTACGA
201 AATCGATGCG GCGGCGAGTG ATTTGCAGAT TTACCAAACC GCCTACAAGG
251 GCGATGCAAC GCCGCCTGAA TGAAGGCGATG GGAAAGCAGG CAGGACGGGC
301 TGCCTTATAA AAACCCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
401 CAACCGCCTG AAAATCGGTA TGCGCCTGCA AACCgAssCG TCCGTGATTT
451 ACGGCATGGG TGCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
501 CGCGACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC
551 GATTGCGCTG CCC..

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

15
1 MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
51 DAFSGNPEGQ FFPDSYEIDA GSDLOIYQT AYKAMQRRLN EAWESRQDGL
101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFNRLKIGM RLQTXSVIY
151 GMGAAYKGKI RKADLRRDTP YNTYTRGGLP PTPIALP..

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

20
25
30
35
1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTTGA CCGTGTCGGC
51 AGCCGTTTTTC GCCGCGCTGC TTTTGTTC TAAGGATAAC GGCAGGGCAT
101 ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCGTCGGT CGGCAGGAAA
151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTCAGCGGC
451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
501 CAGTGATTTG CAGATTTACC AAACCGCTA CAAGGCGATG CAACGCCGCC
551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCCT
601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
651 AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
801 CACCTACAG CGCGGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA
851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT
901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT
951 GACCGAACAC AATGCCGCGG TCCGCAAATA TATTTTGAAA AAATAA

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

40
1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
51 LAEDRIVFSR HVLTAAYVL GVHNLHTGT YRLPSEVSAW DILQMRGGR
101 PDSVTVQIIE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
151 NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRRLEAWES RQDGLPYKNP
201 YEMLIMASLV EKETGHEADR DHVASVFNRL KIGMRLQTD PSVIYGMGAA
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301 FVSKMDGTGL SQFSDLTEH NAAVRKYILK K*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein encoded by yceg gene (accession P44270) of *H.influenzae*

ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

50
55
ORF7 1 MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG 55
+ G+ V+ IEG F RK ++ P + K SNE++ A ++ +
yceg 102 LNSGKEVQFNVKWIEGKTFKDWRKDLENAPHLVQTLKDKSNEEIFALLDLPDIGQNLELK 161
ORF7 56 NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMAXLV 115
N EG +PD+Y +DL++ + + + M++ LN+AW R + LP NPYEMLI+A +V
yceg 162 NVEGWLYPDTYNTYTPKSTDLELLKRSARMKKALNKAWNERDEDLPLANPYEMLILASIV 221
ORF7 116 EKETGHEAXXDHVASVFNRLKIGMRLQTXSVIYGMGAAYKGKIRKADLRRDTPYNTYT 175
EKETG VASVF+NRLK M+LQT +VIYGMG Y G IRK DL TPYNTY
yceg 222 EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV 281

WO 99/24578

PCT/IB98/01665

-75-

ORF7 176 RGGLPPTPIALP 187
 GLPPTPIA+P
 yceg 282 IDGLPPTPIAMP 293

The complete length YCEG protein has sequence:

```

5      1 MKKFLIAILL LILILAGVAS FSYYKMTEFV KTPVNVQADE LLTIERGTTS
      51 SKLATLFEQE KLIADGKLLP YLLKLKPELN KIKAGTYSLE NVKTVQDLLD
     101 LLNSGKEVQF NVKWIEGKTF KDWRKDLENA PHLVQTLKDK SNEEIFALLD
     151 LPDIGQNLEL KNVEGWLYPD TYNYPKSTD LELLKRSAER MKKALNKAWN
     201 ERDEDLPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLKAKMKLQT
     251 DPTVIYGMGE NYNGNIRKRD LETKTPYNTY VIDGLPPTPI AMPSESSLQA
    10 301 VANPEKTDYF YFVADGSGGH KFTRNLNEHN KAVQEYLRWY RSQKNAK
  
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N.*

15 *meningitidis*:

```

                                     10      20      30
orff7.pep                               MRGGRPDSVTVQIIEGSRFSHMRKVIDATP
                                     |||
orff7a      AAYVLGVHNRRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDATP
20      70      80      90      100     110     120

                                     40      50      60      70      80      90
orff7.pep    DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRNLN
25      |||
orff7a      DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAMQRRNLN
      130     140     150     160     170     180

                                     100     110     120     130     140     150
orff7.pep    EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDHVASVFVNRLKIGMRLQTXSVIY
30      |||
orff7a      EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSVIY
      190     200     210     220     230     240

                                     160     170     180
orff7.pep    GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALP
35      |||
orff7a      GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLYFVSKM
      250     260     270     280     290     300

orff7a      DGTGLSQFSHDLTEHNAAVRKYILKKX
40      310     320     330
  
```

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

```

      1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTATCGGC
      51 AGCCGTTTTT GCCGCGCTGC TTTTCGTCCC TAAAGACAAC GGCAGGGCAT
45  101 ACAGGATTAA AATTGCCAAA AACCAGGGTA TTTCGTCTGGT CGGCAGGAAA
      151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
      201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGACTGC
      251 CTTGCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
      301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
50  351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGAACACGAC ACCAAAGGCT
      401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCCCTGATGC CTTAGCGGGC
      451 AATCCTGAAG GGCAGTTTTT CCCCAGACAG TACGAAATCG ATGCGGGCGG
      501 CAGCGATTTA CGGATTTACC AAATCGCCTA CAAGGCGATG CAACGCCGAC
      551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCCT
55  601 TATGAAATGC TGATTATGGC GAGCCTGATC GAAAAGGAAA CAGGGCATGA
      651 AGCCGACCGC GACCATGTCT CTTCCGTCTT CGTCAACCGC CTGAAAATCG
      701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
      751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
      801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATCGCG CTGCCCCGCA
60  851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
      901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATTT
      951 GACCGAACAC AACGCCGCCG TTCGCAAATA TATTTTGAAA AAATAA
  
```

WO 99/24578

PCT/IB98/01665

-76-

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

```

1  MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
51  LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
101 PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
5  151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLEAWES RQDGLPYKNP
201 YEMLIMASLI EKETGHEADR DHVASVFNRL KIGMRLQTD PSVIYGMGAA
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301 FVSKMDGTGL SQFSDHDLTEH NAAVRKYILK K*

```

A leader peptide is underlined.

10 ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

```

10
    10      20      30      40      50      60
orf7a.pep  MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK LAEDRIVFSR
15  orf7-1   MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK LAEDRIVFSR
    10      20      30      40      50      60

    70      80      90     100     110     120
orf7a.pep  HVLTAAYVLGVHNRLHTGT YRLPSEVSAW DILQKMRGGR PDSVTVQIIEGSRFSHMRKV
20  orf7-1   HVLTAAYVLGVHNRLHTGT YRLPSEVSAW DILQKMRGGR PDSVTVQIIEGSRFSHMRKV
    70      80      90     100     110     120

    130     140     150     160     170     180
orf7a.pep  IDATPDIEHDTKGWSNEKLM AEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
25  orf7-1   IDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAM
    130     140     150     160     170     180

    190     200     210     220     230     240
orf7a.pep  QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTD
30  orf7-1   QRRLEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFNRLKIGMRLQTD
    190     200     210     220     230     240

    250     260     270     280     290     300
orf7a.pep  PSVIYGMGAAYKGKIRKADLRRDTPYNTYT RGGLPPTPIALPGKAALDAAAHPSGEKYLY
35  orf7-1   PSVIYGMGAAYKGKIRKADLRRDTPYNTYT RGGLPPTPIALPGKAALDAAAHPSGEKYLY
    250     260     270     280     290     300

    310     320     330
orf7a.pep  FVSKMDGTGLSQFSDHDLTEH NAAVRKYILKKX
40  orf7-1   FVSKMDGTGLSQFSDHDLTEH NAAVRKYILKKX
    310     320     330
45

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N. gonorrhoeae*:

```

50  orf7      MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQ 60
    orf7ng    MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQ 60

    orf7      FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMASLVEKETG 120
55  orf7ng    FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEKETG 120

    orf7      HEAXXDHVASVFNRLKIGMRLQTXSVIYGMGAAYKGKIRKADLRRDTPYNTYT RGGLP 180
60  orf7ng    HEADRDHVASVFNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRRDTPYNTYTGGGLP 180

    orf7      PTPIALP 187

```

		10	20	30	40	50	60
	orf7-1.pep	KLLKWSAVFLTVSAAVF	AALLFV	PKDN	GRAYRIKIAKNQGISSVGRKLAEDRIVFSRHVL		
40	orf7ng-1				YRIKIAKNQGISSVGRKLAEDRIVFSRHVL		
					10	20	30
		70	80	90	100	110	120
	orf7-1.pep	TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPD	SVTVQII	EGSRFSHMRKVIDA			
45	orf7ng-1	TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPD	SVTVQII	EGSRFSHMRKVIDA			
		40	50	60	70	80	90
		130	140	150	160	170	180
50	orf7-1.pep	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR					
	orf7ng-1	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR					
		100	110	120	130	140	150
55		190	200	210	220	230	240
	orf7-1.pep	LNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFVNRLKIGMRLQTDPSV					
	orf7ng-1	LNEAWAGRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSV					
		160	170	180	190	200	210
60		250	260	270	280	290	300
	orf7-1.pep	IYGMGAAYK GKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVS					

WO 99/24578

PCT/IB98/01665

-78-

```

      |||
orf7ng-1  IYGMGAAYKKGKIRKADLRDTPYNTYTGGGLPPTRIALPGKAAMDAAAHPSGEKYLYFVS
           220      230      240      250      260      270
5
      310      320      330
orf7-1.pep KMDGTGLSQFSHDLTEHNAAVRKYILKKX
           |||
orf7ng-1  KMDGTGLSQFSHDLTEHNAAVRKYILKKX
           280      290

```

In addition, ORF7ng-1 shows significant homology with a hypothetical *E.coli* protein:

sp|P28306|YCEG_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG_ECOLI SW: P28306 but
has 97 additional C-terminal residues [Escherichia coli] Length = 340
Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
Identities = 20/87 (22%), Positives = 40/87 (45%)

Query: 10 GISSVGRKLAEDRIVFSRHVLTAAYVLGVHNLHTGTYRLPSEVSAWDILQKMRGGRPD 69
G ++G +L D+I+ V + + GTYR +++ ++L+ + G+
Sbjct: 49 GRLALGEQLYADKIINRPRVFQWLLRIEPLSHFKAGTYRFTPMQTVREMLKLLESKEA 108

Query: 70 SVTVQIIIEGSRFSHMRKVIDATPDIGH 96
++++EG R S K + P I H
Sbjct: 109 QFPLRLVEGMRLSDYKQLREAPYIKH 135

Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
Identities = 84/155 (54%), Positives = 111/155 (71%)

Query: 120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRNEAWAGRQDGLPYKNPYEMLIMASLIEK 179
EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK
Sbjct: 158 EGWFWPDTWMTANTTDVALLKRAHKKMVKAVDSAWEGRADGLPYKDKNQLVTMASIIEK 217

Query: 180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKKGKIRKADLRDTPYNTYTGG 239
ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT
Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277

Query: 240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274
GLPP IA PG ++ AAAHP+ YLYFV+ G
Sbjct: 278 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 6

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>:

```

1  CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT
51 ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG
101 GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC
151 GAACGCGCAC GGCTTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC
50 201 GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGCAGGCG GGAACGGCTC
251 TGGCAACCTA TATGCTGATG TTGGAACGCA CAAAATCCCC CGAAGTCGCC
301 GAACGCGCCT TGGAAATGGC CGTGTCGCTG AACGCGTTTG AACAGGCGGA
351 AATGATTTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC
401 AAAAACGGGC GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG
55 451 CATCTGGACG GACGGGAAGA AGTGCTGGCT CAGGCGGACG AAGGACAG

```

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

```

1  ..REFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFRRKQ QRYSEEEIKN
51 ERARLAAVGE RVNQIFLLG GETALQKGQA GTALATYMLM LERTKSPEVA
101 ERALEMAVSL NAFEQAEMIIY QKWRQIEPIP GKAQKRAGWL RNVLRERGNQ

```

151 HLDGREEVLA QADEGQ

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

	1	ATGTTACCTA	ACCGTTTCAA	AATGTTAACT	GTGTTGACGG	CAACCTTGAT
	51	TGCCGGACAG	GTATCTGCCG	CCGGAGGCGG	TGCGGGGGAT	ATGAAACAGC
5	101	CGAAGGAAGT	CGGAAAGGTT	TTCAGAAAGC	AGCAGCGTTA	CAGCGAGGAA
	151	GAAATCAAAA	ACGAACGCGC	ACGGCTTGCG	GCAGTGGGCG	AGCGGGTTAA
	201	TCAGATATTT	ACGTTGCTGG	GAGGGGAAAC	CGCCTTGCAA	AAGGGGCAGG
	251	CGGGAACGGC	TCTGGCAACC	TATATGCTGA	TGTTGGAACG	CACAAAATCC
	301	CCCGAAGTCG	CCGAACGCGC	CTTGGAATG	GCCGTGTGCG	TGAACGCGTT
10	351	TGAACAGGCG	GAAATGATTT	ATCAGAAATG	GCGGCAGATT	GAGCCTATAC
	401	CGGGTAAGGC	GCAAAAACGG	GCGGGGTGGC	TGCGGAACGT	GCTGAGGGAA
	451	AGAGGAAATC	AGCATCTGGA	CGGACTGGAA	GAAGTGCTGG	CTCAGGCGGA
	501	CGAAGGACAG	AACCGCAGGG	TGTTTTTTATT	GTGGGCACAA	GCCGCCGTGC
	551	AACAGGACGG	GTTGGCGCAA	AAAGCATCGA	AAGCGGTTCTG	CCGCGCGGCG
15	601	TTGAAATATG	AACATCTGCC	CGAAGCGGCG	GTTGCCGATG	TGGTGTTCAG
	651	CGTACAGGGA	CGCGAAAAGG	AAAAGGCAAT	CGGAGCTTTG	CAGCGTTTGG
	701	CGAAGCTCGA	TACGGAAATA	TTGCCCCCA	CTTTAATGAC	GTTGCGTCTG
	751	ACTGCACGCA	AATATCCCGA	AATACTCGAC	GGCTTTTTTCG	AGCAGACAGA
	801	CACCCAAAC	CTTTCGGCCG	TCTGGCAGGA	AATGGAAATT	ATGAATCTGG
20	851	TTTCCCTGCA	CAGGCTGGAT	GATGCCTATG	CGCGTTTGAA	CGTGCTGTTG
	901	GAACGCAATC	CGAATGCAGA	CCTGTATATT	CAGGCAGCGA	TATTGGCGGC
	951	AAACCGAAAA	GAAGTGCTT	CCGTTATCGA	CGGCTACGCC	GAAAAGGCAT
	1001	ACGGCAGGGG	GACGGAGGAA	CAGCGGAGCA	GGGCGGCGCT	AACGGCGGCG
	1051	ATGATGTATG	CCGACCGCAG	GGATTACGCC	AAAGTCAGGC	AGTGGCTGAA
25	1101	AAAAGTATCC	GCGCCGGAAT	ACCTGTTCTGA	CAAAGGTGTG	CTGGCGGCTG
	1151	CGGCGGCTGT	CGAGTTGGAC	GGCGGCAGGG	CGGCTTTGCG	GCAGATCGGC
	1201	AGGGTGCGGA	AACTTCCCGA	ACAGCAGGGG	CGGTATTTTA	CGGCAGACAA
	1251	TTTGTCCAAA	ATACAGATGC	TCGCCCTGTC	GAAGCTGCCC	GATAAACGGG
	1301	AGGCTTTGAG	GGGGTTGGAC	AAGATTATCG	AAAAACCGCC	TGCCGGCAGT
30	1351	AATACAGAGT	TACAGGCAGA	GGCATTGGTA	CAGCGGTCAG	TTGTTTACGA
	1401	TCGGCTTGGC	AAGCGGAAAA	AAATGATTTT	AGATCTTGAA	AGGGCGTTCA
	1451	GGCTTGACAC	CGATAACGCT	CAGATTATGA	ATAATCTGGG	CTACAGCCTG
	1501	CTGACCGATT	CCAAACGTTT	GGACGAAGGT	TTCGCCCTGC	TTCAGACGGC
	1551	ATACCAAATC	AACCCGGACG	ATACCGCTGT	CAACGACAGC	ATAGGCTGGG
35	1601	CGTATTACCT	GAAAGGCGAC	GCGGAAAGCG	CGCTGCCGTA	TCTGCGGTAT
	1651	TCGTTTGAAA	ACGACCCCGA	GCCCGAAGTT	GCCGCCCAT	TGGGCGAAGT
	1701	GTTGTGGGCA	TTGGGCGAAC	GCGATCAGGC	GGTTGACGTA	TGGACGCAGG
	1751	CGGCACACCT	TACGGGAGAC	AAGAAAATAT	GGCGGGAAAC	GCTCAAACGT
	1801	CACGGCATCG	CATTGCCCCA	ACCTTCCCGA	AAACCTCGGA	AATAA

40 This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:

	1	MLPNRFKMLT	VLTATLIAGO	VSAAGGGAGD	MKQPKVEVGKV	FRKQQRYSEE
	51	EIKNERARLA	AVGERVNOIF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIYQKWROI	EPiPGKAQKR	AGWLRNVLRE
	151	RGNQHLDGLE	EVLAQADEGQ	NRRVFLLLAQ	AAVQQDGLAQ	KASKAVRRAA
45	201	LKYEHLPEAA	VADVVFsvQG	REKEKAIGAL	ORLAKLDTEI	LPPTLMTLRL
	251	TARKYPEILD	GFFEQTDTQN	LSAVWQEMEI	MNLVSLHRLD	DAYARLNVLL
	301	ERNPNADLYI	QAAILAANRK	EGASVIDGYA	EKAYGRGTEE	QRSRAALTA
	351	MMYADRRDYA	KVRQWLKKVS	APEYLFdKGv	LAAAAVELD	GGRAALRQIG
	401	RVRKLPEQQG	RYFTADNLSK	IQMLALSKLP	DKREALRGLD	KIIEKPPAGS
	451	NTELQAEALV	QRSVVYDRLG	KRKKMISDLE	RAFRLAPDNA	QIMNNLGYSL
50	501	LTDSKRlDEG	FALLQTAYQI	NPDDTAVNDS	IGWAYYLKGD	AESALPYLRY
	551	SFENDPEPEV	AAHLGEVLWA	LGERDQAVDV	WTQAAHLTGD	KKIWRETLKR
	601	HGIALPQPSR	KPRK*			

Computer analysis of this amino acid sequence gave the following results:

55 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of *N. meningitidis*:

```

                                10      20      30      40      50
60 orf9.pep      RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEGVKQRYSEEEIKNERARLA
      || :|:|:|:|:|:|:|:|:| | | | | | | | | | | | | | | | | | | | | | | |
      orf9a      MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEGVKQRYSEEEIKNERARLA

```

WO 99/24578

PCT/IB98/01665

-80-

		10	20	30	40	50	
		60	70	80	90	100	110
5	orf9.pep	AVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA					
	orf9a	AVGERVNQIFTLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA					
		60	70	80	90	100	110
10	orf9.pep	120	130	140	150	160	
		EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGREEVLAQADEGQ					
	orf9a	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGLEEXLAQADEXQNRVFLLLAQ					
		120	130	140	150	160	170
15	orf9a	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTEI					
		180	190	200	210	220	230

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

	1	ATGTTACCCG	CCCGTTTCAC	CATTTTATCT	GTGCTCGCGG	CAGCCCTGCT
	51	TGCCGGGCAG	GCGTATGCCG	CCGGCGCGGC	GGATGCGAAG	CCGCCGAAGG
20	101	AAGTCGGAAA	GGTTTTCAGA	AAGCAGCAGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAGCGGG	TTAATCAGAT
	201	ATTTACGTTG	CTGGGANGGG	AAACCGCCTT	GCAAAAGGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCNCTGAACG	CGTTTGAACA
25	351	GGCGGAAATG	ATTTATCAGA	AATGGCGGCA	GATTGAGCCT	ATACCGGGTA
	401	AGGCGCAAAA	ACGGGCGGGG	TGGCTGCGGA	ACGTGCTGAG	GGAAAGAGGA
	451	AATCAGCATC	TAGACGGACT	GGAAGAANTG	CTGGCTCAGG	CGGACGAANG
	501	ACAGAACCGC	AGGGTGTTTT	TATTGTTGGC	ACAAGCCGCC	GTGCAACAGG
	551	ACGGGTTGGC	GCAAAAAGCA	TCGAAAGCGG	TTCGCCGCGC	GGCGTTGAGA
30	601	TATGAACATC	TGCCC GAAGC	GGCGGTTGCC	GATGTGGTGT	TCAGCGTACA
	651	GGNACGCGAA	AAGGAAAAGG	CAATCGGAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAATATC	CCGAAATACT	CGACGGCTTT	TTGAGCAGA	CAGACACCCA
	801	AAACCTTTCG	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
35	851	TGCACAGGCT	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACGC
	901	AATCCGAATG	CAGACCTGTA	TATTCAGGCA	GCGATATTGG	CGGCAAACCG
	951	AAAAGAANGT	GCTTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCAGGGCGG	CAATGACGGC	GGCGATGATA
	1051	TATGCCGACC	GAAGGGATTA	CACCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
40	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	TGTGCTGGCG	GCTGCGGCGG
	1151	CTGTGAGTTC	GGACNGCGGC	AGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
	1251	CAAATAACAG	ATGTTGCCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAGGCTT
	1301	TGAGGGGGTT	GGACAAGATT	ATCGAAAAAC	CGCCTGCCCG	CAGTAATACA
45	1351	GAGTTACAGG	CAGAGGCATT	GGTACAGCGG	TCAGTTGTTT	ACGATCGGCT
	1401	TGGCAAGCGG	AAAAAATGA	TTTCAGATCT	TGAAAGGGCG	TTCAGGCTTG
	1451	CACCCGATAA	CGCTCAGATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
	1501	GATTCCAAAC	GTTTGGACGA	AGGCTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CTGTCAACGA	CAGCATAGGC	TGGGCGTATT
50	1601	ACCTGAAANG	CGACGCGGAA	AGCGCGCTGC	CGTATCTGCG	GTATTCGTTT
	1651	GAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTTGTG
	1701	GGCATTGGGC	GAACGCGATC	AGGCGGTTGA	CGTATGGACG	CAGGCGGCAC
	1751	ACCTTACGGG	AGACAAGAAA	ATATGGCGGG	AAACGCTCAA	ACGTCACGGC
	1801	ATCGCATTGC	CCCAACCTTC	CCGAAAACCT	CGGAAATAA	

55 This encodes a protein having amino acid sequence <SEQ ID 44>:

	1	MLPARFTILS	VLAAALLAGQ	AYAAGAADAK	PPKEVGKVFR	KQORYSEEEI
	51	KNERARLAAV	GERVNQIFTL	LGXETALQKG	QAGTALATYM	IMLERTKSPE
	101	VAERALEMAV	SLNAFEQAEM	IYQKWRQIEP	IPGKAQKRAG	WLRNVLRRER
60	151	NQHL DGLEEX	LAQADEXQNR	RVFLLLAQAA	VQQDGLAQKA	SKAVRRAALR
	201	YEHLPEAAVA	DVVFSVQXRE	KEKAIGALQR	LAKLDTEILP	PTLMTLRLTA
	251	RKYPEILDGF	FEQTDTONLS	AVWQEMEIMN	LVSLHRLDDA	YARLNVLLER
	301	NPNADLYIQA	AILAANRKEK	ASVIDGYAEK	AYGRGTGEQR	GRAAMTAAMI
	351	YADRRDYTKV	RQWLKKVSAP	EYLFDKGVLA	AAA AVELDXG	RAALRQIGRV
	401	RKLPEQQGRY	FTADNLSKIQ	MFALSKLPDK	REALRGLDKI	IEKPPAGSNT
65	451	ELQAEALVQR	SVVYDRLGKR	KKMISDLERA	FRLAPDNAQI	MNNLGYSLLS
	501	DSKRLDEGFA	LLQTAYQINP	DDTAVNDSIG	WAYYLKXDAE	SALPYLRYSF
	551	ENDPEPEVAA	HLGEVLWALG	ERDQAVDVWT	QAAHLTGDKK	IWRETLKRHG

WO 99/24578

PCT/IB98/01665

-81-

601 IALPQPSRKPK*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVFRKQORYSEEEIKNERARLA
	orf9-1	10 20 30 40 50 60	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFRKQORYSEEEIKNERARLA
10	orf9a.pep	60 70 80 90 100 110	AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
	orf9-1	70 80 90 100 110 120	AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
15	orf9a.pep	120 130 140 150 160 170	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRVFLLLAQ
	orf9-1	130 140 150 160 170 180	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRVFLLLAQ
20	orf9a.pep	180 190 200 210 220 230	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTEI
	orf9-1	190 200 210 220 230 240	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVFSVQGREKEKAIGALQRLAKLDTEI
25	orf9a.pep	240 250 260 270 280 290	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
	orf9-1	250 260 270 280 290 300	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
30	orf9a.pep	300 310 320 330 340 350	ERNPNADLYIQAAAILAANRKEKXASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYT
	orf9-1	310 320 330 340 350 360	ERNPNADLYIQAAAILAANRKEGASVIDGYAEKAYGRGTEEQRSRAALTAAMMYADRRDYA
35	orf9a.pep	360 370 380 390 400 410	KVRQWLKKVSAPEYLFDKGVLAAAAVELDXGRAALRQIGRVKRLPEQQGRYFTADNLSK
	orf9-1	370 380 390 400 410 420	KVRQWLKKVSAPEYLFDKGVLAAAAVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK
40	orf9a.pep	420 430 440 450 460 470	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTTELQAEALVQSVVYDRLGKRKKMISDLE
	orf9-1	430 440 450 460 470 480	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTTELQAEALVQSVVYDRLGKRKKMISDLE
45	orf9a.pep	480 490 500 510 520 530	RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
	orf9-1	490 500 510 520 530 540	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
50	orf9a.pep	540 550 560 570 580 590	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
	orf9-1	550 560 570 580 590 600	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
55	orf9a.pep	600 610	HGIALPQPSRKPKX
	orf9-1	610	HGIALPQPSRKPKX

WO 99/24578

PCT/IB98/01665

-82-

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N. gonorrhoeae*:

5	Orf9	REFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVERKQORYSEEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVLPKEVGKVLKRRYSEEEIKNERAR	58
10	orf9	LAAVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
	orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
	orf9	QAEMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPAQSDYVHQPMIFLLL	178

- 15 The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

20	1	MIMLPARFTI	LSVLAAALLA	GQAYAAGAAD	VELPKEVGKV	LRKHRRYSEE
	51	EIKNERARLA	AVGERVNRVF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIQKWRQI	EPIPGEAQKP	AGWLRNVLKE
	151	GGNPHLDRLE	EVPAQSDYVH	QPMIFLLLVQ	AAVQHGGVAQ	KPSKAVRPAA
	201	YNYEVLPETA	GADAVFCVQG	PQYEKAIQSF	PPCGRNPQTE	NIAPPFNELF
	251	RPTARPISPK	LLQRFRTTEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

- 25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

30	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCCTCGCAG	CAGCCCTGCT
	51	TGCCGGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGAAA	GGTTTTAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
	201	GTTTACGCTG	TTGGGCGGTG	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
35	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	GGCGGAAATG	ATTTATCAGA	AATGgcggca	gatcgagcct	ataCcggtg
	401	aggcgcaaaa	accgGcgggG	tggctgcgga	acgtattgaa	ggaagggGga
	451	aaTCAGCATC	TGGAcgggtt	gaaagaggTG	CtggcgcaAT	cggacgatGT
	501	GCAAAAACgc	aggatATTTT	TGCTGCTGGT	GCAAGCCGCC	GTGCagcagg
	551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTCCGcgtgc	GGcgttgaAG
40	601	TATGAACATC	TGCCcgaagc	ggcggTTGCC	GATGcggTGT	TCGGCGTACA
	651	GGGACGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAATATC	CCGAAATACT	CGACGGCTTT	TTGAGCAGA	CAGACACCCA
	801	AAACCTTTTCG	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
	851	TGCGTAAGCC	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACAC
45	901	AACCCGAATG	CAAACCTGTA	TATTCAGGCG	GCGATATTGG	CGGCAAACCG
	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCaggcgcg	cAATgacggc	GGCGATGATA
	1051	TATGCCGACC	GCAGGGATTA	CGCCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	CGTGCTGGCG	GCTGCGGCGG
	1151	CTGCCGAATT	GGACGGAGGC	CGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
50	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
	1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAAGCCC
	1301	TGATCGGGCT	GAACAACATC	ATCGCCAAAC	TTTCGGCGGC	GGGAAGCACG
	1351	GAACCTTTTG	CGGAAGCATT	GGCACAGCGT	TCCATTATTT	ACGaacAGTT
	1401	cggCAAACGG	GGAAAAATGA	TTGCCGACCT	tgaAACcgcg	CTCAAACTTA
	1451	CGCCCGATAA	TGCACAAATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
55	1501	GATTCCAAAC	GTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CCGTTAACGA	CAGCATAGGC	TGGGCGTATT
	1601	ACCTGAAAGG	CGACgcggaa	AGCGCGCTGC	CGTATCTGcg	gtattcgttt
	1651	gAAAACGACC	CCGAGCCCGA	AGTTGCCCGC	CATTTGGGCG	AAGTGTGTGT

WO 99/24578

PCT/IB98/01665

-83-

1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
 1751 ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA
 1801 ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA

This encodes a protein having amino acid sequence <SEQ ID 48>:

5 1 MLPARFTILS VLAAALLAGO AYAAGAADVE LPKEVGKVLK KHRRYSEEEI
 51 KNERARLAAV GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE
 101 VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG
 151 NQHL DGLKEV LAQSDDVQKR RIFLLLVQAA VQGGGVAQKA SKAVRRAALK
 201 YEHLPEAAVA DAVFGVQGRE KEKAIEALQR LAKLDTEILP PTLMTLRLTA
 10 251 RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLRKPDDA YARLNVLLEH
 301 NPNANLYIQA AILANRKEG ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI
 351 YADRRDYAKV RQWLKKVSAP EYLFDKGVLA AAAAAELDGG RAALRQIGRV
 401 RKLPEQQGRY FTADNLSKIQ MLALSKLPDK REALIGLNNI IAKLSAAGST
 451 EPLAEALAQR SIIYEQFGKR GKMIADLETA LKLTPDNAQI MNNLGYSLLS
 15 501 DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLKGD AE SALPYLRYSF
 551 ENDPEPEVAA HLGEVLWALG ERDQAVDVT QAAHLRGDKK IWRETLKRYG
 601 IALPEPSRKP RK*

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

20	orf9-1.pep	MLPNR	FKMLT	VLTLT	ATLIAG	QVSAAG	GGGAGD	MKQPK	EVGKVL	RKQQR	YSEEEI	KNERAR	LA
	orf9ng-1	MLPAR	FTILSV	LAAALL	AGQAYA	AG--A	ADVELP	KEVGKV	LKHHRR	YSEEEI	KNERAR	LA	
25	orf9-1.pep	AVGER	VNQIF	TLLG	GETALQ	KGQAG	TALATY	MLMLER	TKSPEV	AERALE	MAVSL	NAFEQA	
	orf9ng-1	AVGER	VNRVFT	LTLG	GETALQ	KGQAG	TALATY	MLMLER	TKSPEV	AERALE	MAVSL	NAFEQA	
30	orf9-1.pep	EMIQK	WRQIE	PIPGK	AQKRAG	WLRNV	LRERGN	QHL DGL	EEVLAQ	ADGQNR	RVFLLL	LAQ	
	orf9ng-1	EMIQK	WRQIE	PIPGK	AQKPAG	WLRNV	LKEGGN	QHL DGL	KEVLAQ	SDDVQK	RRIFLL	LVQ	
35	orf9-1.pep	AAVQQ	DGLAQ	KASKA	VRRRA	ALKYEH	LPEAA	AVADV	VFSVQ	GREKEK	AIGALQ	RLAKLD	TEI
	orf9ng-1	AAVQQ	GGVAQ	KASKA	VRRRA	ALKYEH	LPEAA	AVADV	FGVQ	GREKEK	AIEALQ	RLAKLD	TEI
40	orf9-1.pep	LPPTL	MTLRL	TARKY	PEILD	GFFEQ	TDTON	LSAVW	QEMEIM	NLVSL	HRLDD	AYARLN	VLL
	orf9ng-1	LPPTL	MTLRL	TARKY	PEILD	GFFEQ	TDTON	LSAVW	QEMEIM	NLVSL	RKPDD	AYARLN	VLL
45	orf9-1.pep	ERNPN	ADLYI	QAAIL	AANRKE	GAASV	IDGYAE	KAYGR	GTEEQ	RSRAAL	TAAAMY	ADRRDY	A
	orf9ng-1	EHNP	NANLY	IQAAIL	AANRKE	GAASV	IDGYAE	KAYGR	TGEQR	GRAAMT	AAMIYA	ADRRDY	A
50	orf9-1.pep	KVRQW	LKKVS	AP EYL	FDKGV	LAAAA	AVELD	GGRAAL	RQIGRV	RKLPE	QQGRY	FTADNL	SK
	orf9ng-1	KVRQW	LKKVS	AP EYL	FDKGV	LAAAA	AAELD	GGRAAL	RQIGRV	RKLPE	QQGRY	FTADNL	SK
55	orf9-1.pep	IQMLA	LSKLP	DKREAL	RGLDK	IIIEK	PPAGSN	TELQAE	ALVQR	SVVYDR	LGRKKM	ISDLE	
	orf9ng-1	IQMLA	LSKLP	DKREAL	IGLNNI	IAKLS	AAGST	EPLAE	ALAQR	SIIYE	QFGKR	GKMIAD	LE
60	orf9-1.pep	ERDQ	AVDVT	QAAHL	RGDKK	IWRE	TLKRYG						
65	orf9ng-1	ERDQ	AVDVT	QAAHL	RGDKK	IWRE	TLKRYG						

WO 99/24578

PCT/IB98/01665

-84-

5 orf9-1.pep RAERLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLLKGD
 orf9ng-1 TALKLTPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLLKGD
 480 490 500 510 520 530

10 orf9-1.pep AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWR ETLKR
 orf9ng-1 AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWR ETLKR
 540 550 560 570 580 590

15 orf9-1.pep HGIALPQPSRKPRKX
 orf9ng-1 YGIALPEPSRKPRKX
 600 610

In addition, ORF9ng shows significant homology with a hypothetical protein from *P.aeruginosa*:

20 sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION (ORF3)
 >gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259 (X82071) orf3 [Pseudomonas aeruginosa] Length = 576
 Score = 128 bits (318), Expect = 1e-28
 Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

25 Query: 67 VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIYQKWR 126
 +++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
 Sbjct: 53 LYSLLVAELAGQRNRFDIASNYVVQAQKTRDPGVSEAFRIA EYLGADQEALDTSLLWA 112

30 Query: 127 QIEPIPGEAQKPAG-----WLRNVLKEGGNQHL DGLKEVLAQSDDVQKRRI 172
 + P +AQ+ A ++ VL G+ H D L A++D + +
 Sbjct: 113 RSAPDNLD AQR AAIQLARAGRYEESMVYMEKVLNGQG DTHFDLALSA AETDPDTRAGL 172

35 Query: 173 FXXXXXXXXXXXXXXXXXKASKAVRRAALKY EHLPEAAVADAVFGVQGREKEKAIEALQRLA 232
 ++ KY + + A+ Q ++A+ L+ +
 Sbjct: 173 L-----QSF D HLLKKYPNNGQLLFGKALLLQODGRPDEALTLL EDNS 214

40 Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLRKP 287
 E+ P L + L + K P + G E D + + + + LV +
 Sbjct: 215 ASRHEVAPLLLRSLQLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL----LVEQNRL 270

45 Query: 288 DDAYARLNVLL EHNPN-----ANLYIQAAI----- 312
 DDA A L++ P+ A +Y++ +
 Sbjct: 271 DDAKAEFAGLVQQFPDDDDDLRFS LALVCLEAQAWDEARIYLEELVERDSHVDAAHFNLG 330

50 Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371
 LA +K+ A +D YA+ G G + T ++ A R D A R + P+
 Sbjct: 331 RLAE EQKDTARALDEYAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388

55 Query: 372 YLFDKXXXXXXXXXXXXXXXXXRXQIGRVRLPEQQGRYFTADNLSKIQLALSKLPDKR 431
 Y A L I+ ALS +
 Sbjct: 389 Y-----AIQLYLIEAEALSNN DQOE 408

60 Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
 +A + + + E L L RS++ E+ +M DL + PDNA +
 Sbjct: 409 KAWQAIQEGLKQYP-----EDL-NLLYTRSM LAEKRNDLAQMEKDLRFVIAREPDNAMAL 462

65 Query: 492 NNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLLKGDAESALPYLRY SFE 551
 N LGY+L + R E L+ A++NPDD A+ DS+GW Y +G A YLR + +
 Sbjct: 463 NALGYTLADRTTRYGEARELILKAHKLNPDDPAILD SMGWINYRQGLADAERYLRQALQ 522

70 Query: 552 NDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWR ETLKR 598
 P+ EVA AHLGEVLWA G + A +W + + D + R T+KR
 Sbjct: 523 RYPDHEVA AHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
 Score = 81.5 bits (198), Expect = 1e-14
 Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

Query: 408 GRYFTADNL-SKIQLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQ----- 459
 G Y A L K ++LA PDK+E L + +K + + L +

WO 99/24578

PCT/IB98/01665

-85-

Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390
 Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513
 +I+Y+ G L A++L P+N N LGYSLL +R++E L++
 5 Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYLGYSLLLWYGKERVEEAEELIK 450
 Query: 514 TAYQINPDDTAVNDSIGWAYYLLKGD AESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572
 A + +P++ A DS+GW YYLKGD E A+ YL + E +P V H+G+VL +G +
 10 Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKGDYERAMQYLLKALREAYDDPVVNEHVGDVLLKMGYK 510
 Query: 573 DQAVDVWTQAAHLRGDKK 590
 ++A + + +A L + K
 Sbjct: 511 EEARNYERALKLLEEGK 528

15 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 7

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

20 1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
 51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
 101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
 151 TGGGCGATTA TCGTTTAAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
 201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
 25 251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
 301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CG_aCTGGGCG
 351 GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
 401 GCATTGTTCG CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTGAT
 451 TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCCTGCCC ATCATTATGG
 501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GACCGACCCG
 30 551 ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTGTT
 601 CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACCAACCTCC
 651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
 701 GCCCAAGGCG AAGTCGTTTC CTAA

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

35 1 .NLYAGPQTTs VIANIADNLQ LAKDYGKVHW FASPLEWLLN QLHNIIGNWG
 51 WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQQ
 101 QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLGWI
 151 TDLSRADPYY ILPIIMAATM FAQTYLNPPP TDPMQAKMMK IMPLVFSXXF
 201 FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS *

40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

1 ATGGATTTTA AAAGACTCAC GCGGTTTTTC GCCATCGCGC TGGTGATTAT
 51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
 101 AACAGGCAGC ACAACAACAG GCCGTAACCG CTTCCGCCGA AGCCGCGCTC
 151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
 45 201 TGATGAAAAA AGCGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
 251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGGCGA CGGCAAAGAA
 301 TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
 351 TCTAAAAGGC ATCGGCTTTA GCGCACCAGG AAAACAGTAC AGCTTGGAAG
 401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CGGTCTGAAA
 50 451 ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
 501 CTTGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
 551 ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG TTACTTTACC
 601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
 651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
 55 701 CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC
 751 CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAC AAAGCGTTTG
 801 CGCCGACGGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT
 851 ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA
 901 GCCGAAGCCT CCATCAACCT CTACGCCGGC CCGCAGACCA CATCCGTCAT
 60 951 CGCAAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAAGTAC

WO 99/24578

PCT/IB98/01665

-86-

5
10
15
20
25

```

1001 ACTGGTTCGC CTCCCCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAATGC
1151 GTGCCGCCGC ACCCAAAGTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACAAACAGG GATGATGCAG CTTTACACAG ACGAGAAAAT
1251 CAACCCGCTG GGCGGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401 GCCCATCATT ATGGCGGCAA CGATGTTCGC CCAAACCTAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCGTTGGTT
1501 TTCTCCGTCA TGTTCCTTCT CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA
1601 TCGAAAACA ACGCGCCCAA GGCGAAGTCG TTTCCTAA

```

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

15
20
25

```

1 MDEKRLTAFF AIALVIMIGW EKMFEPTPKPV PAPQQAQQQ AVTASAEAAL
51 APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFILFGDGKE
101 YTYVAQSELL DAQGNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKVYTFTEG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVVYT PEGNFQKVSF SDLDDAKSG KSEAHEYIRKT PTGWLGMIEH
251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
301 AEASINLYAG PQTTSVIANI ADNLQAKDY GKVHWFASPL FWLLNQLHNI
351 IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEKYGDD
401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451 WLGWITDLR ADPYYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
501 FSVMEFFFFPA GLVLYWVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

30
35
40

```

ORF11 2 LYAGPQTTSVIANIADNLQAKDYGVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61
LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K
60K 324 LYAGPKIQSKLKEKELSPGLELTVDYGFLEWFIAPFVLLQHIHSLGNWGSIIIVLTMLIK 383

ORF11 62 AVLYPLTNASYRSMAMRAAPKLQAIKEKYGDDRXXXXXXXXXXLYTDEKINPLGGCLPM 121
+ +PL+ ASYRSM+MRA APKL A+KE++GDDR LY EGINPLGGCLP+
60K 384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443

ORF11 122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLRADPYYILPII MAATMFAQTYLNPPPT 181
L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P
60K 444 LVQMPVFLALYVWLVLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQRNPTPP 503

ORF11 182 DPMQAKMMKIMPLVXXXXXXXXXPAGXVLYWVNNLLTIAQQWHINRSIE 230
DPMQAK+MK+MP++ PAG VLYWVNN L+I+QW+I R IE
60K 504 DPMQAKVMKMMPIIFTFFFLWFAPGLVLYWVNNCLSSISQQWYITRRIE 552

```

45 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N. meningitidis*:

50
55

```

orf11.pep          10      20      30
                   NLYAGPQTTSVIANIADNLQAKDYGVHWF
                   |||
orf11a             IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQAKDYGVHWF
                   280      290      300      310      320      330

                   40      50      60      70      80      90
orf11.pep          FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAPKLQAIKE
                   |||
orf11a             FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAPKLQAIKE
                   340      350      360      370      380      390

```

WO 99/24578

PCT/IB98/01665

-87-

		100	110	120	130	140	150
	orf11.pep	KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI					
	orf11a	KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI					
5		400	410	420	430	440	450
		160	170	180	190	200	210
	orf11.pep	TDLSRADPYYILPII MAATMFAQTYLNPPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLY					
10	orf11a	TDLSRADPYYILPII MAATMFAQTYLNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLY					
		460	470	480	490	500	510
		220	230	240			
	orf11.pep	WVVNNLLTIAQQQWHINRSIEKQRAQGEVVSX					
15	orf11a	WVINNLLTIAQQQWHINRSIEKQRAQGEVVSX					
		520	530	540			

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

	1	ANGGATTTTA	AAAGACTCAC	NGNGTTTTTC	GCCATCGCAC	TGGTGATTAT
20	51	GATCGGATNG	NAAANGATGT	TCCCCACTCC	GAAGCCCGTC	CCCGCGCCCC
	101	AACAGACGGC	ACAACAACAG	GCCGTAANCG	CTTCCGCCGA	AGCCGCGCTC
	151	GCGCCCGNAN	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTCAT
	201	TGATGAAAAA	AGCGGCGACC	TGCGCCGGCT	GACCCGTGCTC	AAATACAAAG
	251	CAACCGGCGA	CNAAAATAAA	CCGTTCATCC	TGTTTGCGCA	CGGCAAANAA
25	301	TACACCTACN	TCGCCCANTC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTAAAAGGC	ATCGGCTTTA	GCGCACCAGAA	AAAACAGTAC	AGCTTGGAAG
	401	GCGACAAAGT	TGAAGTCCGC	CTGAGCGCAC	CTGAAACACG	CGGTCTGAAA
	451	ATCGACAAAG	TTTATACTTT	CACCAAAGGC	AGCTATCTGG	TCAACGTCCG
	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
30	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCGACTTGG	ACGACGATGC	CAANTCCGNN	AAATCCGAGG
	701	CCGAATACAT	CCGCAAAACC	CNGACCGGCT	GGCTCGGCAT	GATTGAACAC
	751	CACTTCATGT	CCACCTGGAT	CCTCCAACCC	AAAGGCGGAC	AAAGCGTTTG
35	801	CGCCGCTGGC	GACTGCNGTA	TNGACATCAA	ACGCCGCAAC	GACAAGCTGT
	851	ACAGCACCAG	CGTCAGCGTG	CCTTTAGCCG	CTATCCAAAA	CGGTGCGAAA
	901	TCCNAAGCCT	CCATCAACCT	CTACGCCGGC	CCACAGACCA	CATCNGTTAT
	951	CGCAAACATC	GCCGACAACC	TGCAACTGGN	CAAAGACTAC	GGCAAAGTAC
	1001	ACTGGTTCGC	CTCCCCCTC	TTTTGGCTTT	TGAACCAACT	GCACAACATC
40	1051	ATCGGCAACT	GGGGCTGGGC	GATTATCGTT	TTAACCATCA	TCGTCAAAGC
	1101	CGTACTGTAT	CCATTGACCA	ACGCCTCTTA	CCGTTGATG	GCGAAAATGC
	1151	GTGCCGCCGC	GCCCAAACCTG	CAAGCCATCA	AAGAGAAATA	CGGCGACGAC
	1201	CGTATGGCGC	AGCAACAAGC	CATGATGCAG	CTTTACACAG	ACGAGAAAAT
	1251	CAACCCGCTG	GGCGGCTGCC	TGCCTATGCT	GTTGCAAATC	CCCGTCTTCA
45	1301	TCGGATTGTA	TTGGGCATTG	TTGCCTCCG	TAGAATTGCG	CCAGGCACCT
	1351	TGGCTGGGTT	GGATTACCGA	CCTCAGCCGC	GCCGACCCNT	ACTACATCCT
	1401	GCCCATCATT	ATGGCGGCAA	CGATGTTTCG	CCAAACCTAT	CTGAACCCGC
	1451	CGCCGACCGA	CCCGATGCAG	GCGAAAATGA	TGAAAATCAT	GCCTTTGGTT
50	1501	NTNTCNNNNA	NGTTCTTCNN	CTTCCCTGCC	GGTCTGGTAT	TGTACTGGGT
	1551	GATCAACAAC	CTCCTGACCA	TCGCCCAGCA	ATGGCACATC	AACCGCAGCA
	1601	TCGAAAACA	ACGCGCCCAA	GGCGAAGTCG	TTTCCTAA	

This encodes a protein having amino acid sequence <SEQ ID 54>:

	1	XDFKRLTXFF	AIALVIMIGX	XXMFPTPKPV	PAPQOTAQQQ	AVXASAEAL
55	51	APXXPITVTT	DTVQAVIDEK	SGDLRRLTLL	KYKATGDXNK	PFILFGDGKX
	101	YTYXAXSELL	DAQNNILKG	IGFSAPKKQY	SLEGDKVEVR	LSAPETRGLK
	151	IDKVYFTFKG	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDAXSG	KSEAEYIRKT	XTGWLGMIEH
	251	HFMSTWILQP	KGGQSVCAAG	DCXXDIKRRN	DKLYSTSVSV	PLAAIQNGAK
	301	SXASINLYAG	PQTTSVIANI	ADNLQLXKDY	GKVHWFASPL	FWLLNQLHNI
60	351	IGNWGWAIIV	LTIIVKAVLY	PLTNASYRSM	AKMRAAAPKL	QAIKEYGDD
	401	RMAQQQAMMQ	LYTDEKINPL	GGCLPMLLQI	PVFIGLYWAL	FASVELRQAP
	451	WLGWITDLR	ADPYIILPII	MAATMFAQTY	LNPPPTDPMQ	AKMMKIMPLV
	501	XSXXFFXFPA	GLVLYWVINN	LLTIAQQQWHI	NRSIEKQRAQ	GEVVS*

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

65

10

20

30

40

50

60

WO 99/24578

PCT/IB98/01665

-88-

5	orf11a.pep	XDFKRLTXFFAIALVIMIGXXXMFPTPKPVPAPQQTAAQQQAVXASAEALAPXXPITVTT
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPAPQQAQQQAVTASAEALAPATPITVTT
10	orf11a.pep	DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFILFGDGKXYTYXAXSELLDAQGNNILKG
	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNNILKG
15	orf11a.pep	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
20	orf11a.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
25	orf11a.pep	XTGWLGMIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK
30	orf11a.pep	SXASINLYAGPQTTSVIANIADNLQLXKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
	orf11-1	AEASINLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
35	orf11a.pep	LTIIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKEKYGDDRMAQQQAMMOLYTDEKINPL
	orf11-1	LTIIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKEKYGDDRMAQQQAMMOLYTDEKINPL
40	orf11a.pep	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY
45	orf11a.pep	LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
50	orf11a.pep	GEVVSX
	orf11-1	GEVVSX

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N. gonorrhoeae*:

65	Orf11	NLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	60

WO 99/24578

PCT/IB98/01665

-89-

	orf11	IIVKAVLYPLTNASYRSMAMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMAMRAAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
5	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	180
10	orf11	PPPTDPMQAKMMKIMPLVFSXXXFFFPAGXVLYWVNNLLTIAQQWHINRSIEKQRAQGE	237
	orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGE	240
	orf11	VVS 240	
15	orf11ng	VVS 243	

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

	1	MAVNLYAGPQ	TTSVIANIAD	NLQLAKDYGK	VHWFASPLFW	LLNQLHNIIG
	51	NWGWAIIVLT	IIVKAVLYPL	TNASYRSMAM	MRAAAPELQT	IKEKYGDDRM
20	101	AQQQAMMQLF	EDEEINPLGG	CLPMLLQIPV	FIGLYWALFA	SVELRQAPWL
	151	GWITDLSRAD	PYYILPIIMA	ATMFAQTYLN	PPPTDPMQAK	MMKIMPLVFS
	201	VMFFFFPAGL	VLYWVNNLL	TIAQQWHINR	SIEKQRAQGE	VVS*

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

	1	ATGGATTTTA	AAAGACTCAC	GGCGTTTTTC	GCCATCGCGC	TGGTGATTAT
25	51	GATCGGCTGG	GAAAAAATGT	TCCCCACCCC	GAAACCCGTC	CCCGCGCCCC
	101	AACAGGCGGC	ACAAAAACAG	GCAGCAACCG	CTTCCGCCGA	AGCCGCGCTC
	151	GCGCCCGCAA	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTTAT
	201	TGATGAAAAA	AGTGGCGACC	TGCGCCGGCT	GACCTGCTC	AAATACAAAG
	251	CAACCGGCGA	CGAAAAACAA	CCGTTCTGTC	TGTTTGGCGA	CGGCAAAGAA
30	301	TACACCTACG	TCGCCCAATC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTGAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	ACCCTCAACG
	401	GCGACACAGT	CGAAGTCCGC	CTGAGCGCGC	CCGAAACCAA	CGGACTGAAA
	451	ATCGACAAAG	TCTATACCTT	TACCAAAGAC	AGCTATCTGG	TCAACGTCCG
	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
35	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCgacTTgg	acgACGATGC	gaaaTccggc	aaATccgagg
	701	ccgaatacaT	CCGCAAAACC	ccgaccggtt	ggctcggcat	gattgaacac
	751	cactttcatgt	ccacctggat	cctccAAcct	aaaggcggcc	aaaacgtttg
40	801	cgcccaggga	gactgccgta	tgcacattaa	aCgccgcaac	gacaagctgt
	851	acagcgcaag	cgtcagcgtg	cctttaaccg	ctatcccaac	ccggggggcca
	901	aaaccgaaaa	tggcggTCAA	CCTGTATGCC	GGTCCGCAAA	CCACATCCGT
	951	TATCGCAAAC	ATCGCcgacA	ACCTGCAACT	GGCAAAAGAC	TACGGTAAAG
	1001	TACACTGGTT	CGCATCGCCG	CTCTTCTGGC	TCCTGAACCA	ACTGCACAAC
45	1051	ATTATCGGCA	ACTGGGGCTG	GGCAATCGTC	GTTTTGACCA	TCATCGTCAA
	1101	AGCCGTACTG	TATCCATTGA	CCAACGcctc	ctACCGTTCTG	ATGGCGAAAA
	1151	TGCGTGccgc	cgcacCcaaA	CTGCAGACCA	TCAAAGAAAA	ATAcgGCGAC
	1201	GACCGTATGG	CGCAACAGCA	AGCGATGATG	CAGCTTTACA	AAgacgAGAA
	1251	AATCAACCCG	CTGGGCGGCT	GTctgcctat	gctgttgCAA	ATCCCCGTCT
50	1301	TCATCGGCTT	GTAATGGGCA	TTGTTTCGCT	CCGTAGAATT	GCGCCAGGCA
	1351	CCTTGGCTGG	GCTGGATTAC	CGACCTCAGC	CGCGCCGACC	CCTACTACAT
	1401	CCTGCCCATC	ATTATGGCGG	CAACGATGTT	CGCCCAAACC	TATCTGAACC
	1451	CGCCGCCGAC	CGACCCGATG	CAGGCGAAAA	TGATGAAAAT	CATGCCGTTG
	1501	GTTTTCTCCG	TCATGTTCTT	CTTCTTCCCT	GCCGGTTTGG	TTCTCTACTG
55	1551	GGTGGTCAAC	AACCTCCTGA	CCATCGCCCA	GCAGTGGCAC	ATCAACCGCA
	1601	GCATCGAAAA	ACAACGCGCC	CAAGGCGAAG	TCGTTTCCTA	A

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

	1	MDFKRLTAFF	AIALVIMIGW	EKMFPPTPKPV	PAPQQAQKQ	AATASAEAL
	51	APATPITVTT	DTVQAVIDEK	SGDLRRLTLL	KYKATGDENK	PFVLFGDGKE
60	101	YTYVAQSELL	DAQNNILKG	IGFSAPKKQY	TLNGDTVEVR	LSAPETNGLK
	151	IDKVYFTFKD	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVYVT	PEGNFQKVSF	SDLDDDAKSG	KSEAEYIRKT	PTGWLGMIEH
	251	HFMSTWILQP	KGGQNVCAQG	DCRIDIKRRN	DKLYSASVSV	PLTAIPTRGP
	301	KPKMAVNLYA	GPQTTSVIAN	IADNLQLAKD	YGKVHWFASP	LEWLLNQLHN

WO 99/24578

PCT/IB98/01665

-90-

351 IIGNWGWAI VLTIIIVKAVL YPLTNASYRS MAKMRAAAPK LQTIKEYGD
 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA
 451 PWLGWITDLS RADPYIILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL
 501 VFSVMFFFFP AGLVLYWVVN NLLTIAQQWH INRSIEKQRA QGEVVS*

5 ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

		10	20	30	40	50	60
	orf11ng-1.pep	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPAPQQAQQAATASAEALAPATPITVTT					
10	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPAPQQAQQAATASAEALAPATPITVTT					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf11ng-1.pep	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNNILKG					
15	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNNILKG					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf11ng-1.pep	IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVYFTTKDSYLVNVRFDIANGSGQTANL					
20	orf11-1	IGFSAPKKQYSLGDKVEVRLSAPETRGLKIDKVYFTTKGSYLVNVRFDIANGSGQTANL					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf11ng-1.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT					
25	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf11ng-1.pep	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGDCRIDIKRRNDKLYSASVSVPLTAIPTRG					
30	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQN-GA					
		250	260	270	280	290	
35		310	320	330	340	350	360
	orf11ng-1.pep	KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAI					
		: :					
40	orf11-1	KAEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAI					
		300	310	320	330	340	350
		370	380	390	400	410	420
	orf11ng-1.pep	VLTIIIVKAVLYPLTNASYRSMAKMRAAAPKLQTIKEYGDDRMAQQQAMMQLYKDEKINP					
45	orf11-1	VLTIIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINP					
		360	370	380	390	400	410
		430	440	450	460	470	480
	orf11ng-1.pep	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQT					
50	orf11-1	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQT					
		420	430	440	450	460	470
		490	500	510	520	530	540
	orf11ng-1.pep	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRA					
55	orf11-1	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRA					
		480	490	500	510	520	530
60		orf11ng-1.pep	QGEVVSX				
	orf11-1	QGEVVSX					
		540					

65 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

WO 99/24578

PCT/IB98/01665

-91-

ID 60IM_PSEPU STANDARD; PRT; 560 AA.
 AC P25754;
 DT 01-MAY-1992 (REL. 22, CREATED)
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE 60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES Init1: 1074 Initn: 1293 Opt: 1103
 Smith-Waterman score: 1406; 41.5% identity in 574 aa overlap

		10	20	30	40
orf11ng-1.pep	MDFKR---	LTAF	FAIALVIMIGW-----	EKMEPT-----	PKPVPA
	:	: :	: :	: :	: :
p25754	MDIKRTILIAALAVVSYVMVLKWNDDY	GQAALPTQNTA	ASTVAPGLPDGVPAGNNGASAD		
		10	20	30	40
		50	60	70	80
orf11ng-1.pep	AATASAEALAPATPIT-----	VTTDTVQ	AVIDEKS	GDRLRLTLLKYKATGDE-NKPF	
	: :	: :	: :	: :	: :
p25754	VPSANAESSPAELAPVALSKDLIRVKT	DVLELAIDPVGGDIVQLNLPKYPRRQDHPNIPF			
		70	80	90	100
		110	120	130	140
orf11ng-1.pep	VLFGDGKEYTYVAQSELLDAQGNILKGIG---	FSAPKKQYTL-NGD---	TVEVRLSAPE		
	:	: :	: :	: :	: :
p25754	QLFDNGGERVYLAQSGLTGTDGPDA-RASGRPLYAAEQKSYQLADGQEQLVVDLKF				
		130	140	150	160
		170	180	190	200
orf11ng-1.pep	TNGLKIDKVYTFKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY				
	:	: :	: :	: :	: :
p25754	DNGVNYIKRFSFKRGEYDLNVSYLIDNQSGQAWNGNMFAQLKR				
		180	190	200	210
		220	230	240	250
orf11ng-1.pep	VGPVVYTPEGNFQKVSFSDLD	DAKSGKSEA	EYIRKTP	TGWLGMIEHHFMSTWILQPKGG	
	: :	: :	: :	: :	: :
p25754	LGAALWTASEPYKKVSMKDID---	KGSLKE-----	NVSGGWV	AWLQHYFVTAWI-PAKSD	
		240	250	260	270
		280	290	300	310
orf11ng-1.pep	QNVCAQGD	CRIDIKRRNDKLYSASVSVPLTAIP	TRGPKPKMAVNLYAGPQTTSVIANIAD		
	: :	: :	: :	: :	: :
p25754	NNV-----	VQTRKDSQGN	YIIGYTGPVISVPA-GGKVETSALLYAGPKIQSKLKELSP		
		290	300	310	320
		330	340	350	360
orf11ng-1.pep	NLQLAKDYGKVHWF-ASPLFWLLNQLHNIIGNWGWAIVVLTII				
	: :	: :	: :	: :	: :
p25754	GLELTVDYGFL-WFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIKGLEFFPLSAASYRSMA				
		340	350	360	370
		380	390	400	410
orf11ng-1.pep	KMRAAAPKLQTIKEKYGDDRMAQQQAMMQLYKDEKINPLGGCLPMLLQIPVFIGLYWALF				
	: :	: :	: :	: :	: :
p25754	RMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPILVQMPVFLALYWVLL				
		400	410	420	430
		440	450	460	470
orf11ng-1.pep	ASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVF				
	:	: :	: :	: :	: :
p25754	ESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQR				
		460	470	480	490
		500	510	520	530
orf11ng-1.pep	SVMFFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGEVVSX				
	: :	: :	: :	: :	: :
p25754	TFFFLWFPAGLVLYWVNNCLSSISQQWYITRRIEAAATKKA				
		520	530	540	550
		560			

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```

1   ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTNG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251 ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

15 This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```

1   ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDL AGQYVEILRH TGGNRYEVXY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```

20  1   ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGTTGT
51  nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GncGCTCTGC TTTCCGCGCT GGGTATTTnG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
25  251 ACCGTTACGA AGTTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```

30  1   ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDL AGQYVEILRH TGGNRYEVFY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.*

35 *meningitidis*:

```

                                10      20      30      40      50
orf13.pep      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
                |||||||
orf13a      MTVWFVA AVAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
                10      20      30      40      50      60
                                60      70      80      90      100     110
orf13.pep      VHAKTAVRKVETDSYQDL DAGQYVEILRH TGGNRYEVXYRGTXWQAQNTGQEELEPGTRA
                ||||| :|||: ||||| ||| |||||
orf13a      VHAKTAVGKVETDSYQDL DAGQYAEILRH AGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                70      80      90      100     110     120
                                120
orf13.pep      LIVRKEGNLLIITHPX
                |||||:|

```

55

1	ATGACTGTAT	GGTTTGGTTGC	CGCTGTTGCC	GTCTTAATCA	TCGAATTATT
51	GACGGGAACG	GTTTATCTTT	TGGTTGTCAG	CGCGGCTTTG	GCGGGTTCGG
101	GCATTGCCCTA	CGGGCTGACT	GGCAGCACGC	CTGCCGCCGT	CTTGACCGCC
151	GCACTGCTTT	CCGCGCTGGG	CATTTGGTTC	GTACATGCCA	AAACCGCCGT
201	GGGAAAAGTT	GAAACGGATT	CATATCAGGA	TTTGGATACC	GGAAAATATG
251	CCGAAATCCT	CCGATACACA	GGCGGCAACC	GTTACGAAGT	TTTTTATCGC
301	GGTACGCACT	GGCAGGCGCA	AAATACGGGG	CAGGAAGTGT	TTGAACCGGG
351	AACGCGCGCC	CTCATCGTCC	GCAAAGAAGG	TAACCTTCTT	ATCATCGCAA
401	ACCCTTAA				

WO 99/24578

PCT/IB98/01665

-94-

This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVSAAL AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYAEILRYT GGNRYEVFYZ
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

10  orf13-1.pep      10      20      30      40      50
      AVLIIELLTGT VYLLVSAAL AGSGIAYGLT GSTPAAVLTX ALLSALGIXF
      |||||
orf13ng 10      20      30      40      50      60
      MTVWFVAAVA VLIIELLTGT VYLLVSAAL AGSGIAYGLT GSTPAAVLTA ALLSALGIWF

15  orf13-1.pep      60      70      80      90     100     110
      VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVFYZRGTHWQAQNTGQEELEPGTRA
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
orf13ng 15      20      30      40      50      60      70      80      90     100     110     120
      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYZRGTHWQAQNTGQEVFEPGTRA

20  orf13-1.pep      120
      LIVRKEGNLLIITHPX
      ||||| :|||
orf13ng 20      30
      LIVRKEGNLLIIANPX
      130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that

25 ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30  1  ATGTwTGATT TCGGTTTrGG CGArCTGGTT TTTGTCGGCA TTATCGCCCT
      51  GATwGtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
      101 GGcTCATCGG CAGGCTGCAA CGCTTTGTCTG GcAGCGTCAA ACAGGAATTT
      151 GACACTCAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
      201 AGCTGCCGcC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
35  251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
      301 CTGCCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
      351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
      401 ATGCCGTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40  1  MXDFGLGELV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
      51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
      101 LPEQRTPADF GVDENGNPXS RCGKHPIRRH FRRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45  1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
      51  GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
      101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
      151 GACACTCAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
      201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
50  251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
      301 CTGCCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
      351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
      401 TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
      451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
      501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG

```

WO 99/24578

PCT/IB98/01665

-95-

551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT
 601 TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTTC GTCCGAAACA
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

5 1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNLP DAANTLSDGI SDVMPSESY ASAETLGDSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT
 201 SLRKQAISRK RDLRPKHSRAK PKLRVRKS*

10 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

 1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
 51 GATTGTCCTC GGCCCCGAAC GCCTGCCCCG ACCGCCGGAC
 101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
 151 GACACGCAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA
 15 201 AGCTGCCGCT GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
 251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCCCGAC AGCGCACGCC TGCTGATTTC GGTGTCGATG AAAACGGCAA
 351 TCCCTTTCCC GATGCGGCAA ACACCCTATT AGACGGCATT TCCGACGTTA
 401 TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
 20 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG
 501 GCGGGAATAC CTGACTGCTT CTGCCGCCG ACCCGTCGTA CAGACCGTCG
 551 AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTCC GCATACCACT
 601 TCGCTGCGTA AACAGGCAAT AAGCCGCAAA CGCGATTTGC GTCCTAAATC
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

25 This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

 1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNFP DAANTLLDGI SDVMPSESY ASAETLGDSG
 30 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT
 201 SLRKQAISRK RDLRPKSRK PKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

		10	20	30	40	50	60
35	orf2.pep	MXDFGLGELVFVGIIALIVLGPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR					
	orf2a	MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR					
		10	20	30	40	50	60
40	orf2.pep	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNXS					
	orf2a	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNFP					
		70	80	90	100	110	120
45	orf2.pep	RCGKHPIRRHFRRYAV					
	orf2a	DAANTLLDGISDVMPSESYASAETLGDSGQTGSTAEPADTDQDRAWREYLTASAAAPVV					
		130	140	150	160	170	180

50 The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

	orf2a.pep	MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
	orf2-1	MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
55	orf2a.pep	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNFP	120
	orf2-1	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNLP	120
60	orf2a.pep	DAANTLLDGISDVMPSESYASAETLGDSGQTGSTAEPADTDQDRAWREYLTASAAAPVV	180

WO 99/24578

PCT/IB98/01665

-96-

orf2-1 DAANTLSDGISDVMPERSYASAETLGDSGQTGSTAETDQDRAWREYLTASAAAPVV 180
 orf2a.pep QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAKPKLRVRKSX 229
 5 orf2-1 QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAKPKLRVRKSX 229

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK
 101 LPEQRTPADF GVDEKGNLSL RYGKHIRRH FRRYAV*

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTCGGCA TTATCGCCCT
 51 GATTGTCCTT GGTCCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC
 15 101 GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT
 151 GACACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCTGA
 201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GATACGGATA
 251 TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCCCGAA AGCGCACGCC tgcgcgatttc gGTGTCGATg AAAacggcaa
 20 351 tcccccttccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA
 401 TGCCGTCTGA ACGTTCCGAT ACTtccgcCG AAACCCTTGG GGACGACAGG
 451 CAAACCGGCA GTACAGCCGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG
 501 GCGGGAATAC CTGactgctt ctgcccgcgc acctgtcgta Cagagggccg
 551 tcgaagtcag ctaTATCGAT ACTGCTGTTG AAacgcctgT tccgcaCacc
 25 601 acttccctgc gcaAACAGGC AATAAACCGC AAACGCGATT TttgtccgaA
 651 ACACCGCGCc aAACCGAAat tgcgcgtcCG TAAATCATAA

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK
 30 101 LPEQRTPADF GVDENGPNLP DTANTVSDGI SDVMPERSD TSAETLGDDR
 151 QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETPVPHT
 201 TSLRKQAINR KRDFCPKHRA KPKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

35 orf2.pep MXDFGLGELVFVGIIALIVLGPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
 orf2ng MFDFGLGELIFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
 40 orf2.pep KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNXS 120
 orf2ng KVKQAFEAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNLPL 120
 orf2.pep RCGKHPIRRHFRRYAV 136
 45 orf2ng RYGKHIRRHFRYAV 136

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

50 orf2-1.pep MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
 orf2ng-1 MFDFGLGELIFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
 55 orf2-1.pep KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNPL
 orf2ng-1 KVKQAFEAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGPNPL

WO 99/24578

PCT/IB98/01665

-97-

```

              70      80      90      100      110      120
              130      140      150      160      170      180
5   orf2-1.pep   DAANTLSDGISDVMPSESYASAETLGDSGQTGSTAEP AETDQDRAWREYLTASAAAPVV
   orf2ng-1      DTANTVSDGISDVMPSESDTSAETLGDDRQTGSTAEP AETDKDRAWREYLTASAAAPVV
              130      140      150      160      170      180
10  orf2-1.pep   Q-TVEVSYIDTAVETVPVHTTSLRKQAISRKRDFRPHRAKPKLRVRKSX
   orf2ng-1      QRAVEVSYIDTAVETVPVHTTSLRKQAINRKRD FCPKRAKPKLRVRKSX
              190      200      210      220      229

```

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined),
 and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein
 of *E.coli*:

```

gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
Score = 56.6 bits (134), Expect = 1e-07
Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)
20  Query: 1  MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
      MF D G  EL+ V II L+VLGP+RLP A +T   I  L+   +V+ EL  +++L+E +
      Sbjct: 1  MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60
25  Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87
      +K+  +A+  +   LK +  +++ +
      Sbjct: 61 DSLKKVEKASLTNLTPELKASMDEL RQA 88

```

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane
 proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above.
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A
 shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results
 of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice,
 whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis
 (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is
 a useful immunogen.

Example 10

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

```

40      1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
      51  CGC.TGCGGG ACACGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT
      101 TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGT TAAA
      151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
      201 CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGCTACTCC
45      251 ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
      301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
      351 GTTTGACAGG TTTAACCACT TCTTTATCTA CACTTAATGC CCCTGCACTC
      401 TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
      451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC

```

WO 99/24578

PCT/IB98/01665

-98-

501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC
 551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
 601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

5 1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT
 101 DYTYPYETT AETTSGLTG LTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
 201 IDVFGTIRNR TEM..

10 Further work revealed the complete nucleotide sequence <SEQ ID 79>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 15 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 201 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 251 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 301 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 20 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCTGCGCG
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAACT GGAATATTTT GCAGTAGACA
 25 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 701 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
 751 AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTTCTCC GATATCCGAC
 801 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
 851 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCAGC AACATAGACA
 901 AGGACAACCT TGA

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPYETT AETTSGLTG LTSLSTLNA PALSRQSDG SGSKSSLGLN
 35 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSKEIK PTEGLMVDFF DIRPYGNHTG NSAPSVEADN
 301 SHEGYGYSDE VVRQHRQGQP *

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

40 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 45 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 50 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCTGCGCG
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAACT GGAATATTTT GCAGTAGACA
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 55 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
 801 AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTTCTCC GATATCCAAC
 851 CATACGGCAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCAGC GACATAGACA
 951 AGGGCAACCT TGA

60 This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK

WO 99/24578

PCT/IB98/01665

-99-

5

```

51  DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPRIYETT AETTSGGLTG LTSLSTLNA PALSRQSDG SGSKSSLGLN
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251 AYKENYALWM GPYKVSCKGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN
301 SHEGYGYSDE AVRRHRQGP *

```

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

10	orf15.pep	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARA AVKMDLQALHGR
	orf15a	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
15	orf15.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDAXXGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
	orf15a	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
20	orf15.pep	130 140 150 160 170 180	LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15a	130 140 150 160 170 180	LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
25	orf15.pep	190 200 210	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM
	orf15a	190 200 210 220 230 240	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

35	orf15a.pep	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
	orf15-1	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
40	orf15a.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
	orf15-1	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
45	orf15a.pep	130 140 150 160 170 180	LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15-1	130 140 150 160 170 180	LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
50	orf15a.pep	190 200 210 220 230 240	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
	orf15-1	190 200 210 220 230 240	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
55	orf15a.pep	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSCKGIKPTGLMVDFS DIQPYGNHMGNSAPSVEADN
	orf15-1	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSCKGIKPTGLMVDFS DIRPYGNHTGNSAPSVEADN
60	orf15a.pep	310 320	SHEGYGYSDEAVRRHRQGPX
	orf15-1	310 320	SHEGYGYSDEVVRQHRQGPX

WO 99/24578

PCT/IB98/01665

-100-

310

320

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

```

      1 ATGCGGGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
      51 CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGCAAACGCT
5  101 TCGCGGTCTGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
      151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
      201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
      251 TTGATGCACT GATTGCGGGC GAATACATAA ACAGCCCTGC CGTCCGCACC
      301 GATTACACCT ATCCGCGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
10  351 TTTGACGGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
      401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA GGAGCAGTCT GGGCTTAAAT
      451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CCAACCCGCG
      501 CGACACTGCC TTTCTTTCCC ACTTGCTGCA GACCGTATTT TTCCTGCGCG
      551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
15  601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
      651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTT GCAGTAGACA
      701 GAACCAATAA AAAATTGCTC ATCAAACCCA AAACCAATGC GTTTGAAGCT
      751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
      801 AGGAATCAAA CCGACGGAAG GATTGATGGT CGATTTCTCC GATATCCAAC
20  851 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
      901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC AACATAGACA
      951 AGGGCAACCT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

```

      1 MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAAVK
25  51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
      101 DYTYPYRSETT AETTSGLTG LTSLSTLNA PALSRTQSDG SGRSSLGLN
      151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVSP ANADTDVFIN
      201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
      251 AYKENYALWM GPYKVSIGIK PTEGLMVDFS DIQPYGNHTG NSAPSVEADN
30  301 SHEGYGYSDE AVRQHRQGP *

```

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

```

      orf15.pep      MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFVEQELVAASARAAVKDMDLQALHGR      60
35  orf15ng          MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAAVKDMDLQALHGR      60
      orf15.pep      KVALYIATMGDQSGSLTGGRYSIDAXXGEYINSPAVRTDYTYPRYETTAETTSGLTG      120
      orf15ng          KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTG      120
40  orf15.pep      LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF      180
      orf15ng          LTSLSTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF      180
45  orf15.pep      FLRGIDVSPANADTDVFINIDVFGTIRNRTEM      213
      orf15ng          FLRGIDVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL      240

```

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

```

50  orf15-1.pep      MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAAVKDMDLQALHGR
      orf15ng          MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAAVKDMDLQALHGR
      10 20 30 40 50 60
55  orf15-1.pep      KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTG
      orf15ng          KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTG
      70 80 90 100 110 120
60  orf15-1.pep      LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF

```

WO 99/24578

PCT/IB98/01665

-101-

	orfl5ng	 LTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF 130 140 150 160 170 180
5	orfl5-1.pep	190 200 210 220 230 240 FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL orfl5ng FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL 190 200 210 220 230 240
10	orfl5-1.pep	250 260 270 280 290 300 IKPKTNAFEAAAYKENYALWMGPYKVSKGIKPTGLMVDIFS DIRPYGNHTGNSAPSVEADN orfl5ng IKPKTNAFEAAAYKENYALWMGPYKVSKGIKPTGLMVDIFS DIQPYGNHTGNSAPSVEADN 250 260 270 280 290 300
15	orfl5-1.pep	310 320 SHEGYGYSDEVVRQHRQGQPX orfl5ng SHEGYGYSDEAVRQHRQGQPX 310 320

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 11

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

35	1	..GG.CAGCACA AAAACAGGC GGTGGAACGG AAAAACCGTA TTTACGATGA
	51	TGCCGGGTAT GATATTCGGC GTATTCACGG GCGCATTCTC CGCAAAATAT
	101	ATCCCCGCGT TCGGGCTTCA AATTTTCTTC ATCCTGTTT TAACCGCCGT
	151	CGCATTCAAA AACTGCATA CCGACCTCA GACGGCATCC CGCCGCTGC
	201	CCGGACTGCC CrGACTGACT GCGGTTTCCA CACTGTTCCG CACAATGTCC
40	251	AGCTGGGTCG GCATAGGCGG CGGTTCACCT TCCGTCCCCT TCTTAATCCA
	301	CTGCGGCTTC CCCGCCATA AAGCCATCGG CACATCATCC GGCCTTGCCT
	351	GGCCGATTGC ACTCTCCGGC GCAATATCGT ATCTGCTCAA CGGCCTGAAT
	401	ATTGCAGGAT TGCCCGAAGG GTCAGTGGG TTCCTTTACC TGCCCGCCGT
	451	CGCCGTCTC AGCGCGGCAA CCATTGCCTT TGCCCCGCTC GGTGTCAAAA
45	501	CCGCCACAA ACTTTCTTCT GCCAACTCA AAAAATC.TT CGGCATTATG
	551	TTGCTTTTGA TTGCCGAAA AATGCTGTAC AACCTGCTTT AA

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

1	..GQHKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILELTAV
51	AFKTLHTDPQ TASRPLPGLP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH

WO 99/24578

PCT/IB98/01665

-102-

101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPVAV
 151 AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL*

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

5
 10
 15
 20

```

1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGCGGC
51 AGGTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
101 CTGTCGTTTT ATGGGTGCTT GATTTCAGG GTTTGGCACA ACATCCTTAC
151 GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
201 CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
251 CCGTATTTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351 GTTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
401 CATCCCGCCC GCTGCCCGGA CTGCCCCGAC TGACTGCGGT TTCCACACTG
451 TTCGGCACAA TGTCGAGCTG GGTCCGCATA GGCGGCGGTT CACTTTCCGT
501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601 CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCCT
651 TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
701 CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAAAA
751 Tc.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801 GCTTTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

25

```

1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51  AQHLAVGTSF AVMVFTEFSS MLGQHKQAV DWKTVFTMMP GMIFGVFTGA
101 LSAKYIPAFG LQIFFILELT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL
151 FGTMSWVGI GGGSLVFPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
251 XFGIMLLLIA GKMLYNLL*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070)

30 ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

35

```

ORF17 3  HKKQAVNGKTVFTMMPGMIFGVFT-GAFAKYIPAFGLQIF--FILELTAVAFKTLHTDP 59
      HK  +  + V  + P ++  VF  G F  +      +IF  +++L      ++  D
HI0902 72 HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130

ORF17 60  QTASRPLPGLPXLTA VSTLFGTMSWVGIGGGSLVFPFLIHCGFPAHKAIGTSSGLAWPI 119
      Q  ++ L  L  +      L G  SS  GIGGG  VPFL  G      +AIG+S+      +
HI0902 131 QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189

ORF17 120 ALSGAISYLLNGLNIAGLPEGSLGFLYLPVAVLSAATIAFAPLGVXXXXXXXXXXXXXXXXX 179
      +SG  S++++G      +PE  SLG++YLPVAV  ++A  +      + LG
HI0902 190 GISGMFSFIVSGWGNPLMPEYSLGYIYLPVVGITATSFFTSKLGASATAKLPVSTLKKG 249

ORF17 180 FGIMLLLIAGKM 191
      F  +  L+++A  M
HI0902 250 FALFLIVVAINM 261
  
```

45

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N.*

meningitidis:

50

```

orf17.pep                                10      20      30
                                           GQHKKQAVNGKTVFTMMPGMIFGVFTGAFS
                                           |||:|||||:|||||:|||||:|||||:
orf17a    QGLAQHPYAQHLAVGTSFAVMVFTEFSSMLGQHKKQAVDWKTVFTMMPGMVFGVFAGALS
           50      60      70      80      90     100

55
           40      50      60      70      80      90
orf17.pep  AKYIPAFGLQIFFILELTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGTMSWVGIGG
           |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf17a    AKYIPAFGLQIFFILELTAVAFKTLHTDPQTASRPLPGLPGLTA VSTLFGTMSWVGIGG
  
```

WO 99/24578

PCT/IB98/01665

-103-

		110	120	130	140	150	160
		100	110	120	130	140	150
5	orf17.pep	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP <u>PAV</u>					
	orf17a	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP <u>PAV</u>					
		170	180	190	200	210	220
		160	170	180	190		
10	orf17.pep	<u>AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX</u>					
	orf17a	<u>AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX</u>					
		230	240	250	260		

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

15	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGCGGC
	51	AGGTTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTGCTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTCATGG	TCTTCACCGC
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GA CTGGAAAA
20	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCATCCT
	351	GTTTTTAACC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG
	451	TTCGGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACTTTCCGT
25	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAAGGGTCAC	TGGGCTTCCT
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACCGCC	CACAACTTTT	CTTCTGCCAA	ACTCAAAAAA
30	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTAA				

This encodes a protein having amino acid sequence <SEQ ID 90>:

	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
	51	AQHLAVGTSF	AVMVETAFFS	MLGQHKQAV	DWKTVFTMMP	GMVFGVFAGA
35	101	LSAKYIPAFG	LQIFFILELT	AVAFKTLHTD	PQTASRPLPG	LPGLTAVSTL
	151	FGTMSSWVGI	GGGSLVPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
	201	LNGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKK
	251	SFGIMLLLIA	GKMLYNLL*			

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

40		10	20	30	40	50	60
	orf17a.pep	MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF					
	orf17-1	MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF					
45		10	20	30	40	50	60
	orf17a.pep	70	80	90	100	110	120
	orf17-1	AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMVFGVFAGALSAKYIPAFGLQIFFILELT					
50		70	80	90	100	110	120
	orf17a.pep	130	140	150	160	170	180
	orf17-1	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA					
55		130	140	150	160	170	180
	orf17a.pep	190	200	210	220	230	240
	orf17-1	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVLSAATIAFAPLGVKTA					
60		190	200	210	220	230	240
	orf17a.pep	250	260	269			
65	orf17a.pep	HKLSSAKLKKSFGIMLLLIAGKMLYNLLX					

orf17-1.pep MWHWDIILILLAVGSAAGFIAGLFGVGGTTLIVPVVLWVLDLQGLAQHPYAOHLAVGTSF

	orfl7ng-1	MWHWDIILILLAVGSAAGFIAGLFGVGGGTLLIVPVVLWVLDLOGLAQHPYAQHLAVGTSTF 10 20 30 40 50 60
5	orfl7-1.pep	AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFFILFLT 70 80 90 100 110 120
	orfl7ng-1	AVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVFAGALSAKYIPAFGLQIFFFILFLT : : 70 80 90 100 110 120
10	orfl7-1.pep	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFMTSSSWVGIGGGSLVPFLIHCGFPAHKA 130 140 150 160 170 180
	orfl7ng-1	AVAFKTLHTGRQTASRPLPGLPGLTAVSTLFAMSSSWVGIGGGSLVPFLIHCGFPAHKA : : 130 140 150 160 170 180
15	orfl7-1.pep	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA 190 200 210 220 230 240
	orfl7ng-1	IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA : : 190 200 210 220 230 240
20	orfl7-1.pep	HKLSSAKLKXFGIMLLLLIAGKMLYNLLX 250 260 269
	orfl7ng-1	HKLSSAKLKESFGIMLLLLIAGKMLYNLLX : : 250 260

In addition, ORF17ng-1 shows significant homology with a hypothetical *H.influenzae* protein:

30 sp|P44070|Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir||G64015 hypothetical protein
HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae
predicted coding region HI0902 [Haemophilus influenzae]Length = 264
Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
Identities = 15/43 (34%), Positives = 23/43 (53%)

```

35 Query:      55 AVGTSFAVMVFETAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVF 97
      A+GTSFA +V T   S   HK   + W+ +   + P ++  VF
Sbjct:      52 ALGTSFATIVITGIGSAQRHHKLGNIWVQAVRILAPVIMLSVF 94

      Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
40      Identities = 44/114 (38%), Positives = 65/114 (57%)

```

45

```
Query: 150 LFGAMSSWVGIGGGSLVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209
      L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G +
Sbjct: 148 LIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207

Query: 210 PEGSLGFLYLPAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263
      PE SLG++YLPVAV ++A + + LG KL + LK+ F + L+++A M
Sbjct: 208 PEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261
```

50

This analysis, including the homology with the hypothetical *H.influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 12

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 95>:

60

1	..GGAAACGGAT	GGCAGGCAGA	CCCCGAACAT	CCGCTGCTCG	GGCTTTTTTG
51	CGTCAGTAAT	GTATCGATGA	CGCTTGCTTT	TGTCGGAATA	TGTGCGTTGG
101	TGCATTATTG	CTTTTCGGGA	ACGGTTCAAG	TGTTTGTGTT	TGCGGCACTG
151	CTCAAACTTT	ATGCGCTGAA	GCCGGTTTAT	TGGTTCGTGT	TGCAGTTTGT
201	GCTGATGGCG	GTTGCCTATG	TCCACCGCTG	CGGTATAGAC	CGGCAGCCGC
251	CGTCAACGTT	CGGCGGCTCG	CAGCTGCGAC	TCGGCGGGTT	GACGGCAGCG

WO 99/24578

PCT/IB98/01665

-106-

301 TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA
351 A

This corresponds to the amino acid sequence <SEQ ID 96; ORF18>:

5 1 ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
51 LKLYALKPVY WFVLQFVLMA VAYVHRCGID RQPPSTFGGS QLRLGGLTAA
101 LMQVSVLVLL LSEIGR*

Further work revealed the complete nucleotide sequence <SEQ ID 97>:

10 1 ATGATTTTGC TGCATTTGGA TTTTGTCT GCCTTACTGT ATGCGGCGGT
51 TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101 GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
151 GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTTCATCC CCCATTTTAA
201 CCTGACTTTG GGCAGCATAT TTTTTCAT CGGGCATTGG AACCGGAAAA
251 CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGGCTT
301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCT GAATATGTGC
15 351 GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTT GTGTTGCGG
401 CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTGCAG
451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501 GCCGCCGTCA ACGTTCGGCG GCTCGCAGCT GCGACTCGGC GGGTTGACGG
551 CAGCGTTGAT GCAGGTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
20 601 AGATAA

This corresponds to the amino acid sequence <SEQ ID 98; ORF18-1>:

25 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
51 GIWGMTRAAP LFIPHFYLTG SIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSM TLAFVICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
151 FVLMAVAYVH RCGIDRQPPS TFGGSQRLRG GLTAALMQVS VLVLLEIG
201 R*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF18 shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) from strain A of *N.*

30 *meningitidis*:

35 orf18.pep 10 20 30
GNGWQADPEHPLLGLFAVSNVSMTLAFVGI
|||
orf18a TRAAPLFIPHFYLTGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI
60 70 80 90 100 110
40 orf18.pep 40 50 60 70 80 90
CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
|||
orf18a CALVHYCFSTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
120 130 140 150 160 170
45 orf18.pep 100 110
QLRLGGLTAALMQVSVLVLLSEIGRX
|||
orf18a QLRLGGLTAALMQSVLVLLSEIGRX
180 190 200

The complete length ORF18a nucleotide sequence <SEQ ID 99> is:

50 1 ATGATTTTGC TGCATTTGGA TTTTGTCT GCCTTACTGT ATGCGGCGGT
51 TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101 GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
151 GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTTCATCC CCCATTTTAA
201 CCTGACTTTG GGCAGCATAT TTTTTCAT CGGGCATTGG AACCGGAAAA
55 251 CGGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCTCT GCTCGGGCTG
301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCT GAATATGTGC
351 GTTGGTGCAT TATTGCTTTT CGNGAACGGT TCAAGTGTTT GTGTTGCGG
401 CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTGCAG

WO 99/24578

PCT/IB98/01665

-107-

451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
 501 GCCGCCGTCA ACGTTCGGCG GNTCGCAGCT GCGACTCGGC GGGTTGACGG
 551 CAGCGTTGAT GCAGNTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
 601 AGATAA

5 This encodes a protein having amino acid sequence <SEQ ID 100>:

1 MILLHLDFLS ALLYAAVFLE LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
 51 GIWGMTRAAP LFIPHFYLTLS GSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
 101 FAVSNVSM TLAFVGICALVH YCFSXTVQVF VFAALLKLYA LKPVYWFVLQ
 151 FVLMAYVHRCGIDRQPPST FGGSQLRLG GLTAALMQXS VLVLLSEIG
 201 R*

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

		10	20	30	40	50	60
15	orf18a.pep	MILLHLDFLSALLYAAVFLELIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
	orf18-1	MILLHLDFLSALLYAAVFLELIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
		10	20	30	40	50	60
20	orf18a.pep	LFIPHFYLTLSGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGICALVH					
	orf18-1	LFIPHFYLTLSGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGICALVH					
		70	80	90	100	110	120
25	orf18a.pep	YCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGSQLRLG					
	orf18-1	YCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGSQLRLG					
		130	140	150	160	170	180
30	orf18a.pep	GLTAALMQXS VLVLLSEIGRX					
	orf18-1	GLTAALMQXS VLVLLSEIGRX					
		190	200				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N. gonorrhoeae*:

40	orf18.pep	GNGWQADPEHPLLGLFAVSNVSM TLAFVGI	30
	orf18ng	TRAAPLFIPHFYLTLSGSIFFFIGYWNKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGI	115
45	orf18.pep	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGS	90
	orf18ng	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGS	175
	orf18.pep	QLRLGGLTAALMQVS VLVLLSEIGR	116
	orf18ng	QLRLGVLAAMLQVAVTAMLLAEIGR	201

50 The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

1 ATGATTTTGC TGCATTTGGA TTTTGTCT GCCTTACTGt aTGC GGcgt
 51 tttTctgTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
 101 GTATTGCGTT GTGGCTCGGC ATCTCGGTTT TAGGGGTAAA GCTGATGCCG
 151 GGGATGTGGG GAATGACCCG CGCCGCGCCT TTGTTCATCC CCCATTTTTA
 55 201 CCTGACTTTG GGCAGCATAT TTTTTCAT CGGGTATTGG AACCGGAAAA
 251 CAGATGGAAA CGGATGGCAG GCAGACCCG AACATCCGCT GCTCGGGCTT
 301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCT GAATATGTGC
 351 GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTC GTGTTGCGG
 401 CATTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTGCAG
 60 451 TTTGTATTGA TGGCGGttgC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
 501 GCCGCCGTCA ACGTTCGGCG GTTCGCAGCT GCGACTCGGC GTGTTGGCGG

WO 99/24578

PCT/IB98/01665

-108-

551 CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC
601 AGATGA

This encodes a protein having amino acid sequence <SEQ ID 102>:

5 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLG ISVLGVKLMP
51 GMWGMTRAAP LFIPHFYLT L GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSM TL AFVGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
151 FVLMAVAYVH RCGIDRQPPS TFGGSQRLRG VLAAMLMOVA VTAMLLAEIG
201 R*

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

10 orf18-1.pep 10 20 30 40 50 60
orf18ng MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
15 orf18-1.pep 70 80 90 100 110 120
orf18ng LFIPHFYLT LGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGICALVH
20 orf18-1.pep 70 80 90 100 110 120
orf18ng LFIPHFYLT LGSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGICALVH
25 orf18-1.pep 130 140 150 160 170 180
orf18ng YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGSQRLRG
30 orf18-1.pep 190 200
orf18ng GLTAALMQVSVLVLLLSEIGRX
orf18ng VLAAMLMOVAVTAMLLAEIGRX
190 200

35 Based on this analysis, including the presence of several putative transmembrane domains in the
gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

40 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
151 TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC
201 CCTGTTTACC CTCTCCTCGC TCACGGCACA AAGCACCTC GGCACAGGGC
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
45 301 GGCGCGGNCG ...

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

1 MKTPLLKPLL ITS LPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNXXTGRLK NIITTVALEFT LSSLTAQSTL GTGLPFILAM TLMTXXFTIL
101 GAX...

50 Further work revealed the complete nucleotide sequence <SEQ ID 105>:

1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC

WO 99/24578

PCT/IB98/01665

-109-

5
10
15
20
25
30
35
40

```

201 CCTGTTACAC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA
301 GGC GCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
401 ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
451 CTGTTCCAAA TCGTCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
551 ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
701 GTTACTACTT TGCCGCCCAA GACATACACG AACGCATCAG CTCCGCCAC
751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
801 CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
851 CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
901 CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
951 CGACAGTCCC GACATCCGCC ACCTGCGCCG CTTTCTCGAC AACCTCGGCA
1001 GCGTCGACCA GCAGTTCGCG CAACTCCAGC ACAACGGCCT GCAGGCAGAA
1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAGC CGCGTCCGCC
1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CGTCTGTGCA AACCAAATC TGGATTGTCA TCGCCAGTAC
1401 CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501 TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
1551 TGCCTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
1801 CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCC AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
2001 GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGCT CGAACCCTAC
2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
2151 A

```

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

45
50
55

```

1 MKTPLLKPLL ITS LPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNRLTGRLK NIITTVALEFT LSSLTAQSTL GTGLPFIAM TLMTEGFETIL
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFIAM GTVLYSTAIL
151 LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPTA KMLRYFFAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG
301 RAIEGCRQSL RLLSDSNDSP DIRHLRRLD NLGSVDQQFR QLQHNGLQAE
351 NDRMGDTRIA ALETSSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAGT
401 IVEALNLTALG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTYG ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPETEPDDF QTALDTRLGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

60

```

orf19 6 LKPLLITSLP VFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLKNIITT 65
L +I+++PVF +V AA +W +MP +LGIIAGGLVDLDN TGRKN+ T
YHFK 5 LNAKVISTIPVFI AVNIAAVGIWFFDISSQSMPLILGIIAGGLVDLDNRLTGRLKNVFFT 64

```

WO 99/24578

PCT/IB98/01665

-110-

orf19 66 VAFTLSSLTAQSTLGTGLPFILAMTLMXXFTILGA 102
 + F++SS Q +G + +I+ MT++T FT++GA
 YHFK 65 LIAFSISSFIVQLHIGHKPIQYIVLMTVLTFTFTMIGA 101

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
10	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLDNXXTGRLK					
	orf19a	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLDNRLTGRLK					
		10	20	30	40	50	60
15	orf19.pep	NIITTVLFTLSSLTAQSTLGTGLPFILAMTLMXXFTILGAX					
	orf19a	NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMFTFGTIMGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
20	orf19a	TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA					
		130	140	150	160	170	180

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

	1	ATGAAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTG	GGCGAACCCA
25	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCTGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTCCAC	CTCTCCTCAC	TTGTCGCGCA	AAGCACCTC	GGCACAGGTT
	251	TGCCATTCAT	CCTCGCCATG	ACCCTGATGA	CTTTCGGCTT	TACCATCATG
	301	GGCGCGGTG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
30	351	CGCCACCTAC	ACCACACTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTTAT	GATTCTGTGC	GGAACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTTCAAGAAA	ACGTCGCCAA
	501	CGCCTACGAA	GCACTCGGCA	GCTACCTCGA	AGCCAAAGCC	GACTTTTTTCG
	551	ATCCCGACGA	AGCCGAATGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
35	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGCCTT	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATACACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
	801	CCGCATCCAC	CGCCTGCTCG	AAATGCAGGG	ACAAGCCTGC	CGCAACACCG
40	851	CCCAAGCCCT	GCGCGCAAGC	AAAGACTACG	TTTACAGCAA	ACGCCTCGGC
	901	CGCGCCATCG	AAGGCTGCCG	CCAATCGCTG	CGCCTCCTTT	CAGACAGCAA
	951	CGACAATCCC	GACATCCGCC	ACCTGCGCCG	CCTTCTCGAC	AACCTCGGCA
	1001	GCGTCGACCA	GCAGTTCCGC	CAACTCCAGC	ACAACGGCCT	GCAGGCAGAA
	1051	AACGACCGCA	TGGGCGACAC	CCGCATCGCC	GCCCTCGAAA	CCGGCAGCCT
45	1101	CAAAAACACC	TGGCAGGCAA	TCCGTCCGCA	GCTAAACCTC	GAATCAGGCG
	1151	TATTCGCCCA	TGCCGTCCGC	CTGTCCCTTG	TCGTTGCCGC	CGCCTGCACC
	1201	ATCGTCGAAG	CCCTCAACCT	CAACCTCGGC	TACTGGATAC	TACTGACCGC
	1251	CCTTTTTCGTC	TGCCAACCCA	ACTACACCGC	CACCAAAGC	CGCGTCCGCC
	1301	AGCGCATCGC	CGGCACCGTA	CTCGGCGTAA	TCGTCGGCTC	GCTCGTCCCC
50	1351	TACTTTACCC	CCTCCGTCGA	AACCAAATC	TGGATCGTCA	TCGCCAGTAC
	1401	CACCTCTTT	TTTATGACCC	GCACCTACAA	ATACAGCTTC	TCGACATTTT
	1451	TCATCACCAT	TCAAGCCCTG	ACCAGCCTCT	CCCTCGCAGG	GTTGGACGTA
	1501	TACGCCGCCA	TGCCCGTACG	CATCATCGAC	ACCATTATCG	GCGCATCCCT
	1551	TGCCTGGGCG	GCAGTCAGCT	ACCTGTGGCC	AGACTGGAAA	TACCTCACGC
55	1601	TCGAACGCAC	CGCCGCCCTT	GCCGTATGCA	GCAACGGCGC	CTATCTCGAA
	1651	AAAATCACCG	AACGCCTCAA	AAGCGGCGAA	ACCGGCGACG	ACGTCGAATA
	1701	CCGCGCCACC	CGCCGCCCGC	CCCACGAACA	CACCGCCGCC	CTCAGCAGCA
	1751	CCCTTTCCGA	CATGAGCAGC	GAACCGCAA	AATTCGCCGA	CAGCCTGCAA
	1801	CCCGGCTTTA	CCCTGCTCAA	AACCGGCTAC	GCCCTGACCG	GCTACATCTC
60	1851	CGCCCTCGGC	GCATACCGCA	GCGAAATGCA	CGAAGAATGC	AGCCCCGACT
	1901	TTACCGCACA	GTTCCACCTC	GCCGCCGAAC	ACACCGCCCA	CATCTTCCAA
	1951	CACCTGCCCG	AAACCGAACC	CGACGACTTT	CAGACAGCAC	TGGATACACT
	2001	GCGCGGCGAA	CTCGACACCC	TCCGCACCCA	CAGCAGCGGA	ACACAAAGCC
	2051	ACATCCTCCT	CCAACAGCTC	CAACTCATCG	CCCGGCAGCT	CGAACCCTAC
65	2101	TACCGCGCCT	ACCGACAAAT	TCCGCACAGG	CAGCCCCAAA	ACGCAGCCTG
	2151	A				

WO 99/24578

PCT/IB98/01665

-111-

This encodes a protein having amino acid sequence <SEQ ID 108>:

```

1  MKTPPLKPLL ITS LPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51  LDNRLTGRLK NIIATVALFT LSSLVAQSTL GTGLPFILAM TLMTFGFTIM
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI
5  151 LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG
301 RAIEGCRQSL RLLSDSNDNP DIRHLRRLD NLGSVDQQFR QLQHNGLQAE
351 NDRMGDTRIA ALETGSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAACT
10  401 IVEALNLNLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQA TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSNGAYLE
551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKGTY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
15  651 HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*

```

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

```

20  orf19a.pep      10      20      30      40      50      60
    MKTPPLKPLLITS LPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
    orf19-1        10      20      30      40      50      60
    MKTPLLKPLLITS LPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK

25  orf19a.pep      70      80      90     100     110     120
    NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY
    orf19-1        70      80      90     100     110     120
    NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY

30  orf19a.pep      130     140     150     160     170     180
    TTLTYTPETYWLTNPFMILCGTVLYSTAIILFQIILPHRPVQENVANAYEALGSYLEAKA
    orf19-1        130     140     150     160     170     180
    TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA

35  orf19a.pep      190     200     210     220     230     240
    DFFDPDEAEWIGNRHIDLAMSNITGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
    orf19-1        190     200     210     220     230     240
    DFFDPDEAAWIGNRHIDLAMSNITGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ

40  orf19a.pep      250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFKNTDIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG
    orf19-1        250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFKNTDIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG

45  orf19a.pep      310     320     330     340     350     360
    RAIEGCRQSLRLLSDSNDNPDIRHLRRLDNLGSVDQQFRQLQHNGLQAEENDRMGDTRIA
    orf19-1        310     320     330     340     350     360
    RAIEGCRQSLRLLSDSNDSPDIRHLRRLDNLGSVDQQFRQLQHNGLQAEENDRMGDTRIA

50  orf19a.pep      370     380     390     400     410     420
    ALETGSLKNTWQAIRPQLNLESGVFRHAVRSLVVAACTIVEALNLNLGYWILLTALFV
    orf19-1        370     380     390     400     410     420
    ALETSSLKNTWQAIRPQLNLESGVFRHAVRSLVVAACTIVEALNLNLGYWILLTALFV

55  orf19a.pep      430     440     450     460     470     480
    CQPNYTATKSRVRQRIAGTVLGIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
    orf19-1        430     440     450     460     470     480
    CQPNYTATKSRVRQRIAGTVLGIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF

60  orf19a.pep      490     500     510     520     530     540
    STFFITIQA TSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL
    orf19-1        490     500     510     520     530     540
    STFFITIQA TSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL

```

WO 99/24578

PCT/IB98/01665

-112-

	orf19-1	STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL	490	500	510	520	530	540
5	orf19a.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSMSSEPAKFADSLQ	550	560	570	580	590	600
	orf19-1	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSMSSEPAKFADSLQ	550	560	570	580	590	600
10	orf19a.pep	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
	orf19-1	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
15	orf19a.pep	QTALDTRLGELDTLRTHSSGTQSHILLQQLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	
	orf19-1	QTALDTRLGELDTLRTHSSGTQSHILLQQLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N. gonorrhoeae*:

25	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKlampFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKlampFVLGIIAGGLVDLDNRLTGRLK	60
30	orf19.pep	NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMXXFTILGAX	103
	orf19ng	NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMFTGFTILGAVGLKYRTFAFGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKlampFV	LGIIAGGLVD
	51	LDNRLTGRLK	NIIATVALEF	LSSLTAQSTL	GTGLPFILAM	TLMTFGFTIL
	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAI
	151	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA	DDFDPEAAW	IGNRHIDLAM
40	201	SNTGVITAFN	QCRSALFYRL	RGKHRHPRTA	KMLRYFQAQ	DIHERISSAH
	251	VDYQEMSEKF	KNTDIIIFRIR	RLLEMQGQAC	RNTAQAIRSG	KDYVYSKRLG
	301	RAIEGCRQSL	RLLSDGNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHSDSPA
	351	NDRMGDTRIA	ALETGSFKNT	*		

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

45	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
50	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCGCA	AAGCACCTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
55	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTCCAAGAAA	GCGTCGCCAA
	501	TGCCTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
	551	ACCCCGATGA	GGCAGCCTGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
60	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
	801	CCGCATCCGC	CGCCTGCTCG	AAATGCAGGG	GCAGGCGTGC	CGCAACACCG
	851	CCCAAGCCAT	CCGGTCGGGC	AAAGACTAcg	tTTACAGCAA	ACGCCTCGGA
	901	CGCGCCATcg	aaggctgCCG	CCAGTCGctg	cgcctCCTTt	cagacggcaA
	951	CGACAGTCCC	GACATCCGCC	ACCTGAGccg	CCTTCTCGAC	AACCTCGgca

WO 99/24578

PCT/IB98/01665

-113-

5
10
15
20
25

```

1001 GCGTcgacca gcagtTCcgc caactCCGAC ACAGcgactC CCCCgcgaa
1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCctcgaaa ccggcagctT
1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCTGCACC
1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC
1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CCTCCGTCGA AACCAAATC TGGATTGTCA TCGCCGGTAC
1401 CACCCTGTTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501 TACGCCGCCA TGCCCGTGCG CATCATcgaC ACCATTATCG GCGCATCCCT
1551 TGCCTGGGCG GCGGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA
1651 AAAATTGCCG AACGCCTCAA AACCGGCGAA ACCGGCGACG ACATAGAATA
1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCGCGAA AATTCGCCGA CAGCCTGCAA
1801 CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT
2001 GCGCGGCGAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC
2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CccgGCAACT CGAACCTAC
2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151 A

```

25 This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

30
35
40

```

1 MKTPLLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNRLTGRK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTEGFTIL
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIR RLLEMQGQAC RNTAQAIRSG KDYVYSKRLG
301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPA
351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVA
401 IVEALNLNLG YWILLTALFV CQPNYTATKS RYQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITIQL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSSGTYLQ
551 KIAERLKTGE TGDDIEYRIT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTYG ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPDMGPDDF QTALDTLRGE LGTLRTRSSG TQSHILLOQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*

```

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

45
50
55
60
65

```

          10      20      30      40      50      60
orf19-1.pep MKTPLLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRK
          |||
orf19ng-1   MKTPLLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRK
          10      20      30      40      50      60

          70      80      90      100     110     120
orf19-1.pep NIITVALFTLSSLTAQSTLGTGLPFILAMTLMTEGFTILGAVGLKYRTFAFGALAVATY
          |||:|||||
orf19ng-1   NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMTEGFTILGAVGLKYRTFAFGALAVATY
          70      80      90      100     110     120

          130     140     150     160     170     180
orf19-1.pep TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA
          |||:|||||
orf19ng-1   TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIILPHRPVQESVANAYEALGGYLEAKA
          130     140     150     160     170     180

          190     200     210     220     230     240
orf19-1.pep DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
          |||
orf19ng-1   DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
          190     200     210     220     230     240

          250     260     270     280     290     300
orf19-1.pep DIHERISSAHVDYQEMSEKFKNTDIIFRIRLLEMQGQACRNTAQALRASKDYVYSKRLG

```

WO 99/24578

PCT/IB98/01665

-114-

5	orf19ng-1	 DIHERISSAHVDYQEMSEKFKNTDIIIFRIRRLLEMQGQACRNTAQAIRSGKDYVYSKRLG 250 260 270 280 290 300
	orf19-1.pep	310 320 330 340 350 360 RAIEGCRQSLRLLSDSNDSPDIRHLRRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA RAIEGCRQSLRLLSDGNDSPDIRHLSRLLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIA 310 320 330 340 350 360
10	orf19-1.pep	370 380 390 400 410 420 ALETSSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNINLGYWILLTALFV ALETGSFKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNINLGYWILLTALFV 370 380 390 400 410 420
	orf19ng-1	370 380 390 400 410 420 ALETGSFKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNINLGYWILLTALFV 370 380 390 400 410 420
15	orf19-1.pep	430 440 450 460 470 480 CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF CQPNYTATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF 430 440 450 460 470 480
	orf19ng-1	430 440 450 460 470 480 CQPNYTATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF 430 440 450 460 470 480
20	orf19-1.pep	490 500 510 520 530 540 STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAAL STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAAL 490 500 510 520 530 540
	orf19ng-1	490 500 510 520 530 540 STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAAL 490 500 510 520 530 540
25	orf19-1.pep	550 560 570 580 590 600 AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ AVCSSGTYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQ 550 560 570 580 590 600
	orf19ng-1	550 560 570 580 590 600 AVCSSGTYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQ 550 560 570 580 590 600
30	orf19-1.pep	610 620 630 640 650 660 PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLLAEHTAHIFQHLPETEPDDF PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLLAEHTAHIFQHLPETEPDDF 610 620 630 640 650 660
	orf19ng-1	610 620 630 640 650 660 PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLLAEHTAHIFQHLPETEPDDF 610 620 630 640 650 660
35	orf19-1.pep	670 680 690 700 710 QTALDTLRGELDTLRTHSSGTQSHILLQQLIARQLEPYRAYRQIPHROPQNAAX QTALDTLRGELDTLRTHSSGTQSHILLQQLIARQLEPYRAYRQIPHROPQNAAX 670 680 690 700 710
	orf19ng-1	670 680 690 700 710 QTALDTLRGELDTLRTHSSGTQSHILLQQLIARQLEPYRAYRQIPHROPQNAAX 670 680 690 700 710

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

50	sp O33369 YOR2_NEIGO HYPOTHETICAL 45.5 KD PROTEIN (ORF2) gnl PID e1154438 (AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417 Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203 Identities = 301/326 (92%), Positives = 306/326 (93%)	
	Query: 307 RQSLRLLSDGNDSPDIRHLSRLLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS 366 RQSLRLLSDGNDSPDIRHLSRLLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS	
55	Sbjct: 1 RQSLRLLSDGNDSDXDIRHLSRLLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS 60	
	Query: 367 FKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNINLGYWILLTALFVCQPNYT 426 FKNTWQAIRPQLNLES VFRHAVRLSLVVAAACTIVEALNINLGYWILLT LFVCQPNYT	
60	Sbjct: 61 FKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNINLGYWILLTRLFVCQPNYT 120	
	Query: 427 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSFSTFFIT 486 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSFSTFFIT	
65	Sbjct: 121 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSFSTFFIT 180	
	Query: 487 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAALAVCSSG 546 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAALAVCSSG	
	Sbjct: 181 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAALAVCSSG 240	

WO 99/24578

PCT/IB98/01665

-115-

Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQPGFTLL 606
 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFAD+ P
 Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADTCNPALPCS 300

5 Query: 607 KTGyALTGYISALGAYRSEMHEECSP 632
 K ALTGYISALG ++ + +P
 Sbjct: 301 KPATALTYGISALGHATAAKCTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology
 10 with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```

15      1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
      51  GCGCGTTTTG GGATTGTGTC GCGATACGGT CATTGCGCGG GCATTCGGCG
     101  CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACG GCCCAACCTG
     151  CTTCGCCGCG TGTTTGC GGA GGGGCGTTT GCCCAAGCGT TTGTGCCGAT
     201  TTTGGCGGAA TACAAGGAAA CGCGTTCAA AGAGGCGG.C GAAGCCTTTA
     20  251  TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
     301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
     351  TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATTT GCTGCGGATT
     401  ACGTTTCCTT ATATATTATT GATTTCCTTG TCTTCATTTG TCGGCTCGGT
     451  ACTCAATTCT TATCATAAGT TCGGCATTCC GGCGTTTACG CCAC.GTTTC
     25  501  TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTTCGAT
     551  CCGCCCGTTA CCGCGCyGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGCA
     601  ACTCGmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
     651  CCAAACtGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
     701  GCGCCTGCgA TTTTgGGCGT GAqCGTGGCG CAGGTTTCTT TGGTGATCAA
     30  751  CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
     801  ACGCCGACCG CATGATGGAG CTGCCCAGCG GCGTGCTGGG GGCGGCACTC
     851  GGTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC
     901  GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGCTgc
     951  TGACGCTGCC GGCGgcGGTC GGACTGGCGG TGTTGTCTGT cCCgCtGGTG
     35  1001  GCGACGCTGT TTATGTACCG CGwATTTACG CTGTTTGACG CGCAGATGAC
     1051  GCAACACGCG CTGATTGCCT ATTCTTTTCG TTTAATCGGC TTAATCATGA
     1101  TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAwAmGCCC
     1151  GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
     1201  CTTTAYCGGC CCACTrrAAC rCaTTCGGAC TTTGCTTGC CATCGGTCTG
     40  1251  GGCGCGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTTGC GCAGACACGG
     1301  TATTTACCAA CCTGG.CAAG GGTTGGGCAG CGTTCTT.AG CAAAATGCT
     1351  GcTCTCGCTC GCCGTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```

45      1  MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPLNL
      51  LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
     101  LGILAAPWVI YVSAPSFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
     151  LNSYHKFGIP AFTPXFLNVS FIVEALFFVP YFDPPVTAXA WAVFVGGILQ
     201  LXFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
     251  TIFASYLQSG SVSWMYYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
     301  EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
     351  QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
     401  FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
     451  SRSP*
  
```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

```

55      1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
      51  GCGCGTTTTG GGATTGTGTC GCGATACGGT CATTGCGCGG GCATTCGGCG
  
```

WO 99/24578

PCT/IB98/01665

-116-

5
10
15
20
25

```

101  CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACCT GCCCAACCTG
151  CTTCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
201  TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGGCG GAGGCTTTTA
251  TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
351  TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTTC CTGCGGATTA
401  CGTTTCCTTA TATATTATTG ATTTCCCTGT CTTCAATTGT CGGCTCGGTA
451  CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT
501  GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCCG TATTTGATC
551  CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTCGGCGG CATTTTGCAA
601  CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
651  CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
701  CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
751  ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
801  CGCCGACCGC ATGATGGAGC TGCCCAGCGG CGTGCTGGGG GCGGCACTCG
851  GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAA CCAAGATACG
901  GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
951  GACGCTGCCG GCGGCGGTTC GACTGGCGGT GTTGTGCTTC CCGCTGGTGG
1001 CGACGCTGTT TATGTACCGC GAATTTACGC TGTTCGACGC GCAGATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT
1101 TAAAGTGTTG GCACCCGGCT TCTATGCGCG GCAAACATC AAAACGCCCCG
1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC
1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
1251 CGCGTGTATC AATGCCGGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
1301 TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTAGCAA AATGCTGCTC
1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC
1401 GTTTGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GGCGGCTTTG
1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAACTGA

```

30 This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

35
40

```

1  MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
51  LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA
101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151 LNSYHKFGIP AFTPTFLNVS FIVEAFFVFP YFDPPVTALA WAVFVGGILO
201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
501 GFRPRHFKRV EN*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

45
50
55
60

```

Orf20 1  MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL LRRVFAEGAF 60
MviN 14  MNLLKSLAAVSSMTMF SRVLGFARDAIVARI FGAAGMATDA FFVAFKLPNL LRRIFAEGAF 73

Orf20 61  AQAFVPILAEYKETRSKEA EAFIRHVAGM LSFVLVIVTALGILAAPWVIYVSAPGFAQD 120
MviN 74  SQAFVPILAEYKSKQGE EATRFVAYVSGLLTLALAVTVAGMLAAPWVIMVTAPGFADT 133

Orf20 121 ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIP AFTPXFLNVS FIVEAFFVFP 180
MviN 134 ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193

Orf20 181 YFDPPVTAXAWAVFVGGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV 240
MviN 194 YFNPPVLALAWAVTVGGVLQVLYQLPYLKKIGMLVLPRINFRDGTAMRVVKQMGPAILGV 253

Orf20 241 SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT 300
MviN 254 SVSQISLIINTIFASFLASGSVSWMYADRLMEFSPGVLGVALGTILLPSLSKSFASGNH 313

```

WO 99/24578

PCT/IB98/01665

-117-

Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 360
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G
 MviN 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKETAFDAAMTQRALIAYSVG 373

5 Orf20 361 LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXXXXCI 420
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+
 MviN 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Orf20 421 NAGLLFYLLRRHGIYQXPXQG 440
 NA LL++ LR+ I+ P G
 MviN 434 NASLLYWQLRKQNIPTPQPG 453

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf20.pep		MNMLGALAKVGSMTVSRVLGFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF					
		:					
orf20a		MNMLGALVKVGSMTVSRVLGFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF					
20		10	20	30	40	50	60
		70	80	90	100	110	120
orf20.pep		AQAFVPILAIEYKETSKEAXEAFIRHVAGMLSEFVLVIVTALGILAAPWVIYVSAPSFAQD					
		:					
orf20a		AQAFVPILAIEYKETSKEATEAFIRHVAGMLSEFVLVIVTALGILAAPWVIYVSAPGFAKD					
25		70	80	90	100	110	120
		130	140	150	160	170	180
orf20.pep		ADKFQLSIDLRLITFPYILLISLSSEFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP					
		:					
orf20a		ADKFQLSIDLRLITFPYILLISLSSEFVGSVLNSYHKFSIPAFTPTFLNVSFIVFALFFVP					
30		130	140	150	160	170	180
		190	200	210	220	230	240
orf20.pep		YFDPFVTAXAWAVFVGGILQLXFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
orf20a		YFDPFVTALAWAVFVGGILQLGFQLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
35		190	200	210	220	230	240
		250	260	270	280	290	300
orf20.pep		SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
		:					
orf20a		SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLGAALGTILLPTLSKHSANQDT					
40		250	260	270	280	290	300
		310	320	330	340	350	360
orf20.pep		EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG					
		:					
orf20a		EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
50		310	320	330	340	350	360
		370	380	390	400	410	420
orf20.pep		LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI					
		:					
orf20a		LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICXQLMNLAFIGPLKHVGLSLAIGLGACI					
55		370	380	390	400	410	420
		430	440	450			
orf20.pep		NAGLLFYLLRRHGIYQXPXQGLGSVLXQKCCSRSPX					
		: ::					
orf20a		NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWLFPDWAHAGGMQKAA					
60		430	440	450	460	470	480

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

65 1 ATGAATATGC TGGGAGCTTT GGTAAGAGTC GGCAGCCTGA CGATGGTGTC
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGC GCATTCGGCG
 101 CAGGCATGGC GACGGATGCG TTCTTTGTGC CGTTCAAACCT GCCCAACCTG

WO 99/24578

PCT/IB98/01665

-118-

5
10
15
20
25

```

151 CTTGCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGACG GAGGCTTTTA
251 TCCGCCATGT GGC GG GATG CTGTCGTTG TACTGGTCAT CGTTACCGCG
301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
351 TGCCAAAGAT GCCGACAAAT TTCAGCTCTC TATCGATTTG CTGCGGATTA
401 CGTTTCCTTA TATCTTATTG ATTTCACTTT CCTCTTTTGT CGGCTCGGTA
451 CTC AATTCCT ATCATAAATT CAGCATTCCT GCGTTTACGC CCACGTTCCCT
501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCTGTGCCG TATTTTCGATC
551 CTCCCGTTAC CGCGCTGGCT TGGGCGGTTT TTGTCGGCGG CATTTTGCAA
601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGTTTTT TGAAACTGCC
651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
701 CGCCTGCGAT TTTGGGCGTG AGCGTGCGC AGATTTCTTT GGTGATCAAC
751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
801 CGCCGACCGC ATGATGGAAC TGCCCGGCGG CGTGCTGGGG GCGGCACTCG
851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCNTGT GCATGCTGCT
951 GACGCTGCCG GCGGCGGTCG GAATGGCGGT GTTGTCTGTT CCGCTGGTGG
1001 CAACCTTGTT TATGTACCGA GAATTCACGC TGTGTGACGC GCAGATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATCATGAT
1101 TAAAGTGTTG GCGCCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
1151 TCAAAATCGC CATCTTCACG CTCATTTGCA CGCAGTTGAT GAACCTTGCC
1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
1251 CGCGTGATC AATGCCGGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
1301 TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTGCAAA AATGCTGCTC
1351 TCGCTCGCCG TGATGGGAGG CGGCCTGTAT GCCGCCCAA TCTGGCTGCC
1401 GTTCGACTGG GCACACGCCG GCGGAATGCA AAAGGCCGCC CGGCTCTTCA
1451 TCCTGATTGC CGTCGGCGGC GGA CTGTATT TCGCATCACT GGCGGCTTTG
1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 118>:

30
35
40

```

1 MNMLGALVKV GSLTMVSRVL GFVRD TVIAR AFGAGMATDA FFVAFKLPNL
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLVIVTA
101 LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151 LNSYHKFSIP AFTPTFLNVS FIVFALFFVP YFDPPTALA WAVEVGGILO
201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN
251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRXCMLLTLP AAVGMAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSEF LIGLIMIKVL APGFYARONI KTPVKIAIFT LICTQLMNLA
401 FIGPLKHVGL SLAIGLGACI NAGLLEYLLR RHGIYQPGKG WAAFLAKMLL
451 SLAVMGGGLY AAQIWLFPDW AHAGGMQKAA RLFILIAVGG GLYFASLAAL
501 GFRPRHFKRVS ES*

```

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

45
50
55
60
65

```

              10      20      30      40      50      60
orf20a.pep  MNMLGALVKVGSGLTMVSRVLGFVRD TVIARA FGA GMA TDA FFVAFKLPNLLRRVFAEGAF
              10      20      30      40      50      60
orf20-1     MNMLGALAKVGSGLTMVSRVLGFVRD TVIARA FGA GMA TDA FFVAFKLPNLLRRVFAEGAF
              70      80      90      100     110     120
orf20a.pep  AQAFVPILA EYKETRSKEATEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFAKD
              70      80      90      100     110     120
orf20-1     AQAFVPILA EYKETRSKEAAEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFAQD
              130     140     150     160     170     180
orf20a.pep  ADKFQLSIDL LRITFPYILLISLSSFVGSVLNSYHKFSIPAFTPTFLNVSFIVFALFFVP
              130     140     150     160     170     180
orf20-1     ADKFQLSIDL LRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPTFLNVSFIVFALFFVP
              190     200     210     220     230     240
orf20a.pep  YFDPPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
              190     200     210     220     230     240
orf20-1     YFDPPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
              250     260     270     280     290     300
orf20a.pep  SVAQISLVINTIFASYLQSGSVSWMYADRM MELPGGVLGAALGTILLPTLSKHSANQDT

```

-119-

Homology with a predicted ORF from *N.gonorrhoeae*

Accession	Gene	Protein	Length
35	orf20.pep	MNMLGALAKVGS LTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 	60
	orf20ng	MNMLGALAKVGS LTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20.pep	AQAFVPILA EYKETRSKEAXEAFIRHVAGMLS FVLVI V TALGILAAPWVIYVSAPSFAQD : : :	120
	orf20ng	AQAFVPILA EYKETRSKEATEAFIRHVAGMLS FVLIVV TALGILAAPWVIYVSAPGFTKD	120
40	orf20.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP : : :	180
	orf20ng	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP	180
45	orf20.pep	YFDP PVTAXAWAVFVGGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV : : :	240
	orf20ng	YFDP PVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV	240
50	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYYADRM MELPSGVLGAALGTILLPTLSKHSANQDT : : :	300
	orf20ng	SVAQISLVINTIFASYLQSGSVSWMYYADRM MELPGGVLGAALGTILLPTLSKHSANQDT	300
55	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQH ALIAYSFG : : :	360
	orf20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQH ALIAYSFG	360
60	orf20.pep	LIGLIMIKVLAPGFYARQNI XXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI : : :	420
	orf20ng	LIGLIMIKVLASGFYARQNI KTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI	420
60	orf20.pep	NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSP 454 : : :	454
	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP 454	454

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

WO 99/24578

PCT/IB98/01665

-120-

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSEVLIVVTA
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGI
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPVPTALA WAVFVGGILO
 5 201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGQG LGQPSWRKCC
 10 451 SRSP*

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
 101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACT GCCCAACCTG
 15 151 CTTCGCCGCG TGTTTGCGGA GGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TATAAGGAAA CGCGTCTTAA AGAGGCGA c gAGGCTTTTA
 251 TCCGCCACGt tgcgggAatg CTGTCGTTTG TGCTGATcgt cGttacCGCG
 301 CTGGGCATAC TTGCCGCGcc tTGGGTGATT TATGTTtccg CgcccGGCTT
 351 TACCAAAGAC GCGGACAAGT TCCAACCTTC CATCAGCCTG CTGCGGATTA
 20 401 CGTTTCCTTA TATATTATTG ATTTCTTTGT CTTCTTTTGT CGGCTCGATA
 451 CTCAATTCCT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT
 501 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCTGCGG TATTTTCGATC
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTTT TTGTCGGCGG TATTTTGCAG
 601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
 25 651 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG
 701 CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta
 801 cgCCGACCGC ATGATGGAGc tgcgcCGGG CGTGCTGGGG GCTGCACTCG
 851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 30 901 GAACAGTTTT CCGCCCTGCT CGACTGGGT TTGCGCCTGT GCATGCTGCT
 951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTCGTTT CCGCTGGTGG
 1001 CGACGCTGTT TATGTACCGA GAATTCACGC TGTTCGACGC ACAAATGACG
 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGT TAATTATGAT
 1101 TAAAGTGTTG GCATCCGGCT TTTATGCGCG GCAAACATC AAAACGCCCCG
 35 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC
 1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCCTGGG
 1251 CGCGTGCATC AACGCCGGAT TGTTGTTCTT CCTGTTGCGC AAACACGGTA
 1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC
 1351 GCGCTCGCCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCCTGCC
 40 1401 GTTCGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGG CAGCTCTGCA
 1451 TCCTGATTGC CGTCGGCGGC GGAAGTATT TCGCATCTCT GGCGGCTTTG
 1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA

This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
 45 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSEVLIVVTA
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGI
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPVPTALA WAVFVGGILO
 201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT
 50 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAAFLAKMLL
 451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAL
 501 GFRPRHFKRV ES*

55 ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

		10	20	30	40	50	60
orf20-1.pep		MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
orf20ng-1		MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
60		10	20	30	40	50	60
		70	80	90	100	110	120
orf20-1.pep		AQAFVPILAE	YKETRSKEA	EAFIRHVAG	MLSFVLVI	VTALGILA	APWVIYVS
orf20ng-1		AQAFVPILAE	YKETRSKEA	EAFIRHVAG	MLSFVLVI	VTALGILA	APWVIYVS
65							

WO 99/24578

PCT/IB98/01665

-121-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf20-1.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPTFLNVSFIVFALFFVP					
	orf20ng-1	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP					
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf20-1.pep	YFDPPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV					
	orf20ng-1	YFDPPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV					
		190	200	210	220	230	240
		250	260	270	280	290	300
15	orf20-1.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
	orf20ng-1	SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
		310	320	330	340	350	360
20	orf20-1.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf20-1.pep	LIGLIMIKVLAPGFYARONIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
30	orf20ng-1	LIGLIMIKVLASGFYARONIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
		370	380	390	400	410	420
		430	440	450	460	470	480
	orf20-1.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
35	orf20ng-1	NAGLLFLLRKHGIYRPGRWAAFLAKMLLALAVMCGGLWAAQACLPEWAHAGGMRKAG					
		430	440	450	460	470	480
		490	500	510			
40	orf20-1.pep	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVESX					
		490	500	510			

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

45	sp P37169 MVIN_SALTY_VIRULENCE_FACTOR_MVIN_pir S40271 mviN protein - Salmonella typhimurium gi 438252 (Z26133) mviB gene product [Salmonella typhimurium] gnl PID d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524	
	Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220	
	Identities = 309/467 (66%), Positives = 368/467 (78%)	
50	Query:	1 MNMLGALAKVGSLTMVSRVLGFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60
		MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF
	Sbjct:	14 MNLLKSLAAVSSMTMFSRVLGFAIDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
55	Query:	61 AQAFVPILAIEYKETSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD 120
		+QAFVPILAIEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF
	Sbjct:	74 SQAFVPILAIEYKSKQGEETRIFFVAYVSGLLTALAVVTVAGMLAAPWVIMVTAPGFADT 133
60	Query:	121 ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP 180
		ADKF L+ LLRITFPYILLISL+S VG+ILN+++F IPAF PTFLNIS I FALF P
	Sbjct:	134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193
	Query:	181 YFDPPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV 240
		YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAILGV
65	Sbjct:	194 YFNPPVLALAWAVTVGGVLQVYQLPYLKKIGMLVLPRINFRDTGAMRVVKQMGPAILGV 253
	Query:	241 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT 300
		SV+QISL+INTIFAS+L SGSVSWMYADR+ME GVLG ALGTILLP+LSK A+ +
70	Sbjct:	254 SVSQISLIINTIFASFLASGSVSWMYADRLMEFPVSGVLGVALGTILLPSLSKSFASGNH 313

WO 99/24578

PCT/IB98/01665

-122-

5 Query: 301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG 360
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G
 Sbjct: 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSFG 373

5 Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIPLKHAGLSLAIGLGACI 420
 LIGLI++KVL GFY+RQ+IKTPVKIAI TLI TQLMNLAFIPLKHAGLSL+IGL AC+
 Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIPLKHAGLSLSIGLAACL 433

10 Query: 421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACLP 467
 NA LL++ LRK I+ P GW VM L+ +P
 Sbjct: 434 NASLLYWQLRKQNIPTPQPGWWMFLMRIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 14/41 (34%), Positives = 23/41 (56%)

15 Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509
 EW+ + + +L ++ G YFA+LA LGF+ + F R
 Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521

20 Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

25 1 atGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT tACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 151 GTcAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGCGGAAA
 30 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTTCAG CAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA tGGACACCAA TCCG..

35 This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFta PASGKIAAIH RGEKRVLQSV VIAVEXNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTP..

40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

1 ATGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 45 151 GTCAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGCGGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTTCAG CAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 50 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATTAT
 501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT
 551 TGACCGAACG CAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
 651 TGCCGGTTTG AGTGGCACGC ACATTCAATT CATCGAGCCG GTCGGCGCGA
 701 ATAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
 801 TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
 901 TCCGTTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT

WO 99/24578

PCT/IB98/01665

-123-

5
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
 1051 ACAACCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC
 1101 CGTCAACGGC GGCACCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCGATACCG ACAGCGCGCA GGCATTGGGT TGCTTGGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

10
 1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVITIGR
 15
 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELEFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

20
 1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
 201 GTTTACCGCG CCNGTTTCAG GCAAATCGC CGCCATCCAT CGCGGCGAAA
 25
 251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACCTAA GCGGCGANGA
 351 ANTNNGNNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
 401 GTCCGTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTGTGAT
 30
 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
 551 TGACCGAGCG TAAATCCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
 651 GGCCGGTTTG AGTGGCACGC ACATTCAATTT CATTGAGCCG GTCGGTGCAA
 701 ACAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
 35
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
 801 TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
 901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 40
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
 1051 ACGACCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGT GGCACCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCGATACCG ACAGCGCGCA AGCATTGGGT TGCTTGGAAT TGGACGAAGA
 45
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
 1301 CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

50
 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PXMKVKEGDA
 51 VKKGQVLFED KXPGVVFTA PVSGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGXEXXX NLIQSGLWTA LRXRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPVVVIKEA XXDFRRXXLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDADNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELEFGWVA PQPDKYSITR
 55
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

60
 orf22.pep 10 20 30 40 50 60
 MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
 orf22a MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED

WO 99/24578

PCT/IB98/01665

-124-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf22.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR					
	orf22a	KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX					
		70	80	90	100	110	120
10	orf22.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP					
	orf22a	NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV					
		130	140	150	160	170	180

The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:

15	orf22a.pep	10	20	30	40	50	60
	orf22-1	10	20	30	40	50	60
20	orf22a.pep	70	80	90	100	110	120
	orf22-1	70	80	90	100	110	120
25	orf22a.pep	130	140	150	160	170	180
	orf22-1	130	140	150	160	170	180
30	orf22a.pep	190	200	210	220	230	240
	orf22-1	190	200	210	220	230	240
35	orf22a.pep	250	260	270	280	290	300
	orf22-1	250	260	270	280	290	300
40	orf22a.pep	310	320	330	340	350	360
	orf22-1	310	320	330	340	350	360
45	orf22a.pep	370	380	390	400	410	420
	orf22-1	370	380	390	400	410	420
50	orf22a.pep	430	440				
	orf22-1	430	440				
55	orf22a.pep						
	orf22-1						
60	orf22a.pep						
	orf22-1						

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

65	1	MIKIKKGLNL	PIAGRPEQVI	YDGPATEVA	LLGEEYVGM	PSMKIKEGEA
	51	VKKGQVLFED	KKNPGVVFTA	PASGKIAAIH	RGEKRVLSV	VI AVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	R NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF

WO 99/24578

PCT/IB98/01665

-125-

151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV
 201 SENAANIETH EFGGPHAPAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFVTGRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
 301 SGSVLNGAIA QGAHDYLGRY HN*

5 Further work identified complete gonococcal gene <SEQ ID 131>:

1 ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTAGT
 10 201 ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTCAC CGTGGCGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATTGA GCAGCGAAAA
 351 AGTGCGCCGC AACCTGATTC AATCAGGCTT ATGGACTGCG CTTGCGACCC
 401 GTCCGTTTCAG CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC
 15 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATCAT
 501 CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC
 551 TGACCGAAGC TAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC
 651 TGCCGCTTG AGTGGCACGC ACATTCATTT CATCGAGCCA GTCGGCGCGA
 20 701 ATAAACCGT GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT
 751 TTGTTCTGTA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CCTTGGGCGG
 801 CCTGCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAGG
 851 TGTCTCAACT TACCGCCGGC GAATTGGTTG ACGCGGACAA CCGCGTGATT
 901 TCCGGTTCGG TATTGAACGG TGCATTGCA CAAGGCGCGC ATGATTATTT
 25 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAATACTC CATCACGCGC
 1051 ACCACTCTCG GCCATTTCCT AAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGC GGCGACCGCG CCATGGTACC GATCGGCACT TATGAGCGCG
 1151 TAATGCCGTT GGACATCCTG CCTACCTTGC TTTTGCGCGA TTTAATCGTC
 30 1201 GGCGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

1 MIKIKKGLNL PIAGRPEQVI YDGPATEVA LLGEEYVGM RPSMKIKEGEA
 35 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYVPEAL AKLSSEKVR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV
 201 SENAANIETH EFGGPHAPAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFVTGRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
 40 301 SGSVLNGAIA QGAHDYLG RYHNQISVIEEG RSKELEFGWVA PQPKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRLKVL ETIEKEG*

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa

45 overlap with ORF22ng:

orf22.pep	MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED	60
orf22ng	MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED	60
orf22.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR	120
orf22ng	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR	120
orf22.pep	NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP	158
orf22ng	NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV	180

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

60	orf22-1.pep	MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
----	-------------	---

WO 99/24578

PCT/IB98/01665

-126-

```

|||||:|||||:|||||:|||||:|||||
orf22ng-1  MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED
           10      20      30      40      50      60
5
           70      80      90     100     110     120
orf22-1.pep KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYAPEALANLSGEEVRR
           |||||:|||||:|||||:|||||:|||||:|||||
orf22ng-1  KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYVPEALAKLSSEKVR
           70      80      90     100     110     120
10
           130     140     150     160     170     180
orf22-1.pep NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV
           |||||:|||||:|||||:|||||:|||||:|||||
orf22ng-1  NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV
           130     140     150     160     170     180
15
           190     200     210     220     230     240
orf22-1.pep LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFI EPVGANKTVWTI
           |||||:|||||:|||||:|||||:|||||:|||||
orf22ng-1  LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFI EPVGANKTVWTI
           190     200     210     220     230     240
20
           250     260     270     280     290     300
orf22-1.pep NYQDVITIGRLFATGRINTERVIALGGSQVNKPRLRLTVLGAKVSQITAGELVDTDNRVI
           |||||:|||||:|||||:|||||:|||||:|||||
orf22ng-1  NYQDVIAIGRLFVTGRINTERVVALGGQLQVNKPRLRLTVLGAKVSQLTAGELVDADNRVI
           250     260     270     280     290     300
25
           310     320     330     340     350     360
orf22-1.pep SGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTLGHFLKNK
           |||||:|||||:|||||:|||||:|||||:|||||
orf22ng-1  SGSVLNGAIAQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTLGHFLKNK
           310     320     330     340     350     360
30
           370     380     390     400     410     420
orf22-1.pep LFKFNTAVNGGDRAMVPIGTYERVMPLDILPTLLRLDLIVGDTDSAQALGCLELDEEDLA
           |||||:|||||:|||||:|||||:|||||:|||||
orf22ng-1  LFKFTTAVNGGDRAMVPIGTYERVMPLDILPTLLRLDLIVGDTDSAQALGCLELDEEDLA
           370     380     390     400     410     420
35
           430     440
orf22-1.pep LCSFVCPGKYEYGPLL R KVLETIEKEGX
           |||||:|||||:|||||:|||||:|||||:|||||
orf22ng-1  LCSFVCPGKYEYGPLL R KVLETIEKEGX
           430     440
40
45

```

Computer analysis of these sequences gave the following results:

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

ORF22 and this 48kDa protein show 72% aa identity in 158aa overlap:

```

50
Orf22  1  MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKKGQVLFED 60
      MI IKKGL+LPIAG P Q +++G + EVA+LGEEY GMRPSMKV+EGD VKKGQVLFED
48kDa  1  MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

orf22  61  KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYAPEALANLSGEEVRR 120
      KKNPGVVFTAPASG + I+RGEKRVLSVVI VE +++I F RY LA+LS E+v++
55
48kDa  61  KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

orf22  121 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158
      NLI+SGLWTA RTRPFSSK+PA+DA P +IFVNAMDTNP
60
48kDa  121 NLIESGLWTAFRTRPFSSKVPALDAIPSSIFVNAMDTNP 158

```

ORF22a also shows homology to the 48kDa *Actinobacillus pleuropneumoniae* protein:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]
Length = 449

Score = 530 bits (1351), Expect = e-150

WO 99/24578

PCT/IB98/01665

-127-

Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

5 Query: 1 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED 60
MI IKKGL+LPIAG P QVI++G + EVA+LGEY GM RP MKV+EGD VKKGQVLFED
Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

10 Query: 61 KKNPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXX 120
KK PGVVFTAP SG + I+RGEKRVLSVVI VEG+++I F RY LA+LS +
Sbjct: 61 KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

15 Query: 121 NLIQSGLWTALRXPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180
NLI+SGLWTA R RPF SK+PA+DA P +IFVNAMDTNPLAADP VV+KE DF+ V
Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

20 Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTV 237
L+RL ++ +++CK A +++P S I F G HPAGL GTHIHF++PVGA K V
Sbjct: 181 LTRLFNGQKPVYLCKDADSNIP LSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240

25 Query: 238 WTINYQDVIAIGRLFATGRLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADN 297
W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQ+TA EL +N
Sbjct: 241 WHLNYQDVIAIGKLETTGELFTDRIISLAGPQVKNPRLVRLGANLSQLTANELNAGEN 300

30 Query: 298 RVISGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELEFGWVAPQPKYSITRTTLGHFL 357
RVISGSVL+GA G DYLG RY Q+SV+ EGR KELEFGW+ P DK+SITRT LGHF
Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELEFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KKNLFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXXXXVGDTDSAQXXXXXXXXXX 417
K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ
Sbjct: 361 K-KLFNFTTAVHGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 418 XXXXXSFVCPGKYEXGPLL R KVLETXEKEG 447
++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A. pleuropneumoniae*:

35 gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus
pleuropneumoniae] Length = 449
Score = 555 bits (1414), Expect = e-157
Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

40 Query: 27 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 86
MI IKKGL+LPIAG P QVI++G + EVA+LGEYVGM RPSMK++EG+ VKKGQVLFED
Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

45 Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 146
KKNPGVVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++
Sbjct: 61 KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

50 Query: 147 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVVIKEAAEDFKRGLLV 206
NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNPLAADP V++KE DFK GL V
Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

55 Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTV 263
L+RL ++ +++CK A +++P S I F G HPAGL GTHIHF++PVGA K V
Sbjct: 181 LTRLFNGQKPVYLCKDADSNIP LSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240

60 Query: 264 WTINYQDVIAIGRLFVTGRLNTERVVALGGLQVKNPRLRLTVLGAKVSQITAGELVDADN 323
W +NYQDVIAIG+LF TG L T+R+++L G QV PRL+RT LGA +SQLTA EL +N
Sbjct: 241 WHLNYQDVIAIGKLETTGELFTDRIISLAGPQVKNPRLVRLGANLSQLTANELNAGEN 300

65 Query: 324 RVISGSVLNGAIAQGAHDYLG RYHNQISVIEEGRSKELEFGWVAPQPKYSITRTTLGHFL 383
RVISGSVL+GA A G DYLG RY Q+SV+ EGR KELEFGW+ P DK+SITRT LGHF
Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELEFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KKNLFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXXXXVGDTDSAQXXXXXXXXXX 443
K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ
Sbjct: 361 K-KLFNFTTAVHGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 444 XXXXXSFVCPGKYEYGP LL R KVLETIEKEG 473
++VCPGK YGP+LR LE IEKEG
Sbjct: 420 DLALCTYVCPGKNNGPMLRAALEKIEKEG 449

70

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

10 Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```

1   .GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
101 TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
15  151 TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
251 TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20  401 GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGGsGC TTTmTTTGSW CakcATCTTT TTTGCCGCAC AGTTTGTGCG
501 ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551 CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTGTT TATCGGTTTT
601 ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
25  651 ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701 ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751 AATATTATTA CGCCGATGAT GAGTTATTTT GGGCTGATTA TGGCGACGGT
801 GrkCmmTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851 TGCCGTATTC CGCTTCTTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
30  901 TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
951 TCCCGCACCT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

1   .AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
35  101 TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151 TLXLXLXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIFG
201 ILICAFINLM IGSASAQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
251 NIITPMMSYF GLIMATVXXY KKDAGVGTLL XMMLPYSAFF LIAWIALFCI
301 WVFVLGLPVG PGAPTFYPAP *

```

40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

1   ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTTT ATTATTTTCA
101 TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
151 GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
45  201 GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
251 CGCATACCGT TAAAAATTTT ACCGGTTTCG CGCCGTTGGG AACGGTGTTG
301 GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCGCG
351 ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAATC ACTACTTTTA
401 TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
50  451 GTCGTCTTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
501 TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT

```

WO 99/24578

PCT/IB98/01665

-129-

5 551 CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
 601 CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
 651 CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
 701 ATTTTGTAC TGAAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
 751 GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
 801 TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTC GTTGCCTTAT
 851 CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
 901 CCTGAAACAG GATTGGTTTC CGGTTCCGGG TTTTAAAT CGATTGTTGT
 951 TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATTGTTTAT GGCCGGGTAA
 10 1001 CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
 1051 ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
 1101 TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
 1151 GGGCGACGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTATATC
 1201 GGTTTTATTT TAATTGTGTC TTTTATCAAT CTGATGATAG GCTCCGCCTC
 15 1251 CGCGCAATGG GCGGTAAGT GCGCGATTTT CGTCCCTATG CTGATGTTGG
 1301 CCGGCTACGC GCGGTAAGT ATTCAAGCCG CTTACCGCAT CCGTGATTCC
 1351 GTTACCAATA TTATTACGCC GATGATGAGT TATTTCCGGC TGATTATGGC
 1401 GACGGTGATC AAATACAAA AAGATGCGGG CGTGGGTACG CTGATTTCTA
 1451 TGATGTTGCC GTATTCCGCT TTCTTCTTGA TTGCGTGGAT TGCCTTATTC
 20 1501 TGCATTTGGG TATTTGTTTT GGGCCTGCCC GTCGGTCCCG GCGCGCCCAC
 1551 ATTCTATCCC GCACCTTAA

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

1 MSQTDTRDQ RFLRTVEWLG NMLPHPVTLE IIFIVLLLIA SAVGAYFGLS
 51 VPDPRPVGAK GRADDGLIYI VSLNADGFI KILHTVKNF TGFAPLGTVL
 25 101 VSLLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
 151 VVLIPLSAII FHSLSGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
 201 QQAAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
 251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
 301 PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
 30 351 MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLFI
 401 GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
 451 VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMLPYSA FFLIAWIALF
 501 CIWVFVGLP VGPGAPTFYP AP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N. meningitidis*:

40	orf12.pep	AXXIIHPXXVVGPEANWFFMVASTFVIALI
	orf12a	AAAFAGVSGGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMVASTFVIALI
		180 190 200 210 220 230
45	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
	orf12a	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
		240 250 260 270 280 290
50	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
	orf12a	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMS
		300 310 320 330 340 350
55	orf12.pep	TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
	orf12a	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
		360 370 380 390 400 410
60	orf12.pep	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY
		220 230 240 250 260 270

WO 99/24578

PCT/IB98/01665

-130-

5	orf12a	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY
		420 430 440 450 460 470
5	orf12.pep	280 290 300 310 320
		KKDAGVGTLLIXMMLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFYPAPX
	orf12a	KKDAGVGTLLISMMLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFYPAPX
		480 490 500 510 520

The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

10	1	ATGAGTCAAA	CCGATACGCA	ACGGGACGGA	CGATTTTAC	GCACAGTCGA
	51	ATGGCTGGGC	AATATGTTGC	CGCACCCGGT	TACGCTTTT	ATTATTTTCA
	101	TTGTGTTATT	GCTGATTGCC	TCTGCCGCCG	GTGCGTATTT	CGGACTATCC
	151	GTCCCCGATC	CGCGCCCTGT	TGGTGCAGAA	GGACGTGCCG	ATGACGGTTT
	201	GATTCACGTT	GTCAGCCTGC	TCGATGCTGA	CGGTTTGATC	AAAATCCTGA
	251	CGCATACCGT	TAAAAATTTT	ACCGGTTTCG	CGCCGTTGGG	AACGGTGTTG
15	301	GTTTCTTTAT	TGGGCGTGGG	GATTGCGGAA	AAATCGGGCT	TGATTTCCGC
	351	ATTAATGCGC	TTATTGCTCA	CAAAATCTCC	ACGCAAATC	ACTACTTTTA
	401	TGGTTGTTTT	TACAGGGATT	TTATCTAATA	CCGCTTCTGA	ATTGGGCTAT
	451	GTCGTCCTAA	TCCCTTTGTC	CGCCATCATC	TTTCATTCCC	TCGGCCGCCA
20	501	TCCGCTTGCC	GGTCTGGCTG	CGGCTTTCGC	CGGCGTTTCG	GGCGGTTATT
	551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTTGGC	AGGCATCACC
	601	CAACAGGCGG	CGCAAATCAT	CCATCCCGAC	TACGTCGTAG	GCCCTGAAGC
	651	CAACTGGTTT	TTTATGGTAG	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
	701	ATTTTGTTAC	TGAAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA
	751	GATTTGTCAC	AAGAAGAAAA	AGACATTCGA	CATTCCAATG	AAATCACGCC
25	801	TTTGGAATAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTTT	GTTGCCTTAT
	851	CCGCCCTATT	GGCTTGGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCCTCAT
	901	CCTGAAACAG	GATTGGTTTC	CGGTTCCCGG	TTTTTAAAT	CAATTGTTGT
	951	TTTTATTTTC	TTGTTGTTTG	CACTGCCGGG	CATTGTTTAT	GGCCGGGTAA
30	1001	CCCGAAGTTT	GCGCGGCGAA	CAGGAAGTCG	TTAATGCGAT	GGCCGAATCG
	1051	ATGAGTACTC	TGGGGCTTTA	TTTGGTCATC	ATCTTTTTTG	CCGCACAGTT
	1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	GCCGTAAAG
	1151	GGGCGACGTT	CTTAAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC
	1201	GGTTTTATTT	TAATTTGTGC	TTTTATCAAT	CTGATGATAG	GCTCCGCCTC
	1251	CGCGCAATGG	GCGGTAAGTG	CGCCGATTTT	CGTCCCTATG	CTGATGTTGG
35	1301	CCGGCTACGC	GCCCGAAGTC	ATTCAAGCCG	CTTACCGCAT	CGGTGATTCC
	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTGCGGC	TGATTATGGC
	1401	GACGGTGATC	AAATACAAAA	AAGATGCGGG	CGTGGGTACG	CTGATTTCTA
	1451	TGATGTTGCC	GTATTCCGCT	TTCTTCTTGA	TTGCGTGGAT	TGCCTTATTC
40	1501	TGCATTTGGG	TATTTGTTTT	GGGCCTGCCC	GTCGGTCCCG	GCGCGCCAC
	1551	ATTCTATCCC	GCACCTTAA			

This encodes a protein having amino acid sequence <SEQ ID 138>:

45	1	MSQTDQORDG	REFLRTVEWLG	NMLPHPVTLE	IIFIVLLLIA	SAAGAYFGLS
	51	VDPDRPVGAK	GRADDGLIHV	VSLLDADGLI	KILTHTVKNF	TGFAPLGTVL
	101	VSLLGVGIAE	KSLGISALMR	LLLTSPRKL	TFMVVFTGI	LSNTASELGY
	151	VLIPLSAII	FHSLGRHPLA	GLAAAFAGVS	GGYSANLFLG	TIDPLLAGIT
	201	QQAQIIHPD	YVVGPEANWF	FMVASTFVIA	LIGYFVTEKI	VEPQLGPYQS
	251	DLSQEEKDIR	HSNEITPLEY	KGLIWAGVVF	VALSALLAWS	IVPADGILRH
50	301	PETGLVSGSP	FLKSIVVEIF	LLFALPGIVY	GRVTRSLRGE	QEVVNAMAES
	351	MSTLGLYLVI	IFFAAQEVAF	FNWTNIGQYI	AVKGATFLKE	VGLGGSVLFI
	401	GFILICAFIN	LMIGSASAQW	AVTAPIFVPM	LMLAGYAPEV	IQAAYRIGDS
	451	VTNIITPMMS	YFGLIMATVI	KYKKDAGVGT	LISMMLPYSA	FFLIAWIALF
	501	CIWVFLGLP	VGPGAPTFYP	AP*		

55 ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

60	orf12a.pep	10 20 30 40 50 60
		MSQTDQORDGRFLRTVEWLG NMLPHPVTLE IIFIVLLLIASAAGAYFGLSVDPDRPVGAK
	orf12-1	10 20 30 40 50 60
		MSQTDQORDGRFLRTVEWLG NMLPHPVTLE IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
65	orf12a.pep	70 80 90 100 110 120
		GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSGLISALMR
	orf12-1	70 80 90 100 110 120
		GRADDGLIYIVSLLNADGFIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSGLISALMR

WO 99/24578

PCT/IB98/01665

-131-

		70	80	90	100	110	120
5	orf12a.pep	130	140	150	160	170	180
	orf12-1	130	140	150	160	170	180
10	orf12a.pep	190	200	210	220	230	240
	orf12-1	190	200	210	220	230	240
15	orf12a.pep	250	260	270	280	290	300
	orf12-1	250	260	270	280	290	300
20	orf12a.pep	310	320	330	340	350	360
	orf12-1	310	320	330	340	350	360
25	orf12a.pep	370	380	390	400	410	420
	orf12-1	370	380	390	400	410	420
30	orf12a.pep	430	440	450	460	470	480
	orf12-1	430	440	450	460	470	480
35	orf12a.pep	490	500	510	520		
	orf12-1	490	500	510	520		

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N. gonorrhoeae*:

50	orf12.pep	AXXIIHPXXVVGPEANWFFMVASTFVIALI	30
	orf12ng	AAAFAGVSGGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALI	232
55	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	90
	orf12ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	292
60	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	150
	orf12ng	PADGILRHPETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAXAESMS	352
65	orf12.pep	TLXLXLXIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM	210
	orf12ng	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGAVFLKKFRLGGSVLFIFIGFILICAFINLM	412
	orf12.pep	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY	270
	orf12ng	IGSASAQWAVTAPIFVPMMLLAGNAPQVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY	472

WO 99/24578

PCT/IB98/01665

-132-

orfl2.pep KKDAGVGTLIXMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFYPAP 320
 |||||:|||||:
 orfl2ng KKDAGVGTLISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTTFYPVP 522

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCTGT TACGCTTTT ATTATTTTCA
     101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATTT CGGACTATCC
     151  GTCCCCGATC CGCGTCCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
     201  GATTCACGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCCTGA
     251  CGCATACCGT TAAAAATTC ACCGTTTCG CGCCGTTGGG AACGGTGTTG
    10  301  GTTCTTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAATCCCC ACGCAAATC ACTACTTTTA
     401  TGGTTGTTT TACAGGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
     451  GTCGTCCTAA TCCCTTTGTC CGCCGTCATC TTTCATTCGC TCGGCCGCCA
    15  501  TCCGCTTGCC GGTGTTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
     551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTTAC TGAAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
    20  751  GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
     801  TTTGGAATAT AAAGGATTAA TTTGGGCAGG CGTGGTGTGT GTTGCCTTAT
     851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTGC CGGTTCCCGG TTTTAAATAT CGATTGTTGT
     951  TTTTATTTTC TTGTTGTTTG CGTGCCGGG CATTGTTTAT GGCCGGATAA
    25 1001  CCCGAAGTTT GCGCGGCGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
     1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
     1101  TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTAAAG
     1151  GGGCGGTGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTATTATC
     1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
    30 1251  CGCGCAATGG GCGGTAATG CGCCGATTTT CGTCCCTATG CTGATGTTGG
     1301  CCGGCTACGC GCGGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
     1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCCGGC TGATTATGGC
     1401  GACGGTAATC AAATACAAA AAGATGCGGG CGTAGGCACG CTGATTTCTA
     1451  TGATGTTGCC GTATTCCGCT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
    35 1501  TGCATTTGGG TATTTGTTTT GGGTCTGCCC GTCGGTCCCG GCACACCCAC
     1551  ATTCTATCCG GTGCCTTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

      1  MSQTDARRSG RFLRTVEWLGNMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
      51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
    40  101  VSLLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
     151  VVLIPLSAVI FHSLSGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
     201  QQAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
     251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
     301  PETGLVAGSP FLKSIVVFIF LLEALPGIVY GRITRSLRGE REVVNAMAES
    45  351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGAVFLKK FRLGGSVLFI
     401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
     451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
     501  CIWVFLGLP VPGTPTTFYP VP*
  
```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50      10      20      30      40      50      60
orfl2-1.pep MSQTDTDQDGRFLRTVEWLGNMLPHPVTLFIIIFIVLLLIASAVGAYFGLSVPDPRPVGAK
      |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl2ng     MSQTDARRSGRFLRTVEWLGNMLPHPVTLFIIIFIVLLLIASAVGAYFGLSVPDPRPVGAK
      10      20      30      40      50      60

55      70      80      90      100     110     120
orfl2-1.pep GRADDGLIYIVSLLNADGFIKILTHTVKNF TGFAPLGTVLVSLLGVGIAEKSLISALMR
      |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl2ng     GRADDGLIHVVSLLDADGLIKILTHTVKNF TGFAPLGTVLVSLLGVGIAEKSLISALMR
      70      80      90      100     110     120

60      130     140     150     160     170     180
orfl2-1.pep LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLSGRHPLAGLAAAFAGVS
      |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
65  orfl2ng     LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLSGRHPLAGLAAAFAGVS
  
```

WO 99/24578

PCT/IB98/01665

-133-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf12-1.pep	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI :					
	orf12ng	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKI :					
		190	200	210	220	230	240
		250	260	270	280	290	300
10	orf12-1.pep	VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH :					
	orf12ng	VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH :					
		250	260	270	280	290	300
		310	320	330	340	350	360
15	orf12-1.pep	PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAEISMSTLGLYLVI : :					
	orf12ng	PETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAEISMSTLGLYLVI : :					
		310	320	330	340	350	360
		370	380	390	400	410	420
20	orf12-1.pep	IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFILICAFINLMIGSASAQW :					
	orf12ng	IFFAAQFVAFFNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFILICAFINLMIGSASAQW :					
25		370	380	390	400	410	420
		430	440	450	460	470	480
	orf12-1.pep	AVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT :					
30	orf12ng	AVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT :					
		430	440	450	460	470	480
		490	500	510	520		
35	orf12-1.pep	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAPX :					
	orf12ng	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTFFYPVPX :					
		490	500	510	520		

In addition, ORF12ng shows significant homology with a hypothetical protein from *E.coli*:

40	sp P46133 YDAH ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION >gi 1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli] Length = 510 Score = 329 bits (835), Expect = 2e-89 Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)	
45	Query: 8	RSGRFLRTVEWLGNNMLPHPVTXXXXXXXXXXASAVGAYFGLSVPDPRPVGAKGRADDGL 67 +SG+ VE +GN +PHP +A+ + FG+S +P D
	Sbjct: 13	QSGKLYGWVERIGNKVPHPFLFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64
50	Query: 68	IHVSLLDADGLIKILHTVKNFTGFAPXXXXXXXXXXIAEKSGLISALMRLLLTKSP 127 + V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + +
	Sbjct: 65	VVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124
55	Query: 128	RKLTFMVVFTGILSNTASELGYYVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187 + ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL
	Sbjct: 125	ARYASYMVLFIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVCGFTANL 184
60	Query: 188	FLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247 + T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG
	Sbjct: 185	LIVTTDVLLSGISTEAAAANFNPQMHVSVIDNWFMASSVVVLTIVGGLITDKIIEPRLGQ 244
65	Query: 248	YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRHPETGLVA 307 +Q + ++ + + S GL AGVV + A +A ++P +GILR P V
	Sbjct: 245	WQGNSDEKLQTLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298
70	Query: 308	GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAEISMSTLGLYLXXXXXXXXXX 367 SPF+K IV I L F + + YG TR++R + ++ + M E M + ++
	Sbjct: 299	PSPFIKGIVPLIILFFFVSLAYGIATRTIRROADLPHLMIEPMKEMAGFIVMVFLAQF 358
	Query: 368	XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFILICAFINLMIGSASAQWAVTAPIF 427 NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF

5
10
15
20
25
30
35

```

1  MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
51  LRVIGCGGVA CLPDFQONVG EADFAVVPDD AAARAVIEV DADDAVCAQK
101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
151 QFGFFRVGGA SFVITAQAGI DDALCDCCLTA DAAGFAVFAF VADGQMVFEG
201 NVQPAVETGF FHGISVSSVF GAAQYSAMA SRSASIPVFS ATEMRTAAIF
251 PAASRHMPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLEFA
301 AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
351 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*

```

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

15
20
25

```

1  ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
51  GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTTATTGTG ATGATTTTGA
101 TGCCGAAGTC GGGCAGCTTC GGTTCGGCT ATGCGTCGCT GGCGGCTTTG
151 TCGTTCGGCG CGCTGATGAT TCGCTGTTA GACGTGTCGT CAAATATGGC
201 GATGCAGCCG TTAAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
301 GTGGCGGCGA TTCTGCCGTT TGTGTTTGGC TATATCGGTT TGGCGAACAC
351 CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG
401 GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCCG
501 GAATCAGGAA AAAGCCAACCT GGATCGCACT CTAAAAA.CC GCGC..

```

25 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

```

1  ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL
51  SFGALMIALL DVSSNMAMQP FKMMVGDMDVN EEQKXYAYGI QSFLANTGAV
101 VAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK
151 EYXPETYARY HGIDVAANQE KANWIALLKX A..

```

30 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

35
40
45
50
55

```

1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
51  AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
101 CCTTTACCTT GCAAAGCTCG CAAATGAGCC GCATTTTTC AAGCCTAGGC
151 GCAGACCCGC ACAATTTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG
201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
251 CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
351 CTATGCGTCG CTGGCGGCTT TGTCGTTCCG CGCGCTGATG ATTGCGCTGT
401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501 CTTAGCAAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
601 GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
651 GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAACC TACGCCGTT
701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
751 CTCTTGAAAA CCGCGCCTAA GCGGTTTTGG ACGGTTACTT TGGTGCAATT
801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
951 GGTGATTGTG TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
1101 CATCGCTTGG GCGGGCATT AACTTATCC GCTGACGATT GTGACCAACG
1151 CCTTGTCGGG CAAGCATATG GGCACCTACT TGGGCTTGTT TAACGGCTCT
1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
1251 TATGCTGGGC GGCTTGCAAG CCACTATGTT CTTGGTAGGG GCGTCGTCC
1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTTGA TTAAAGAAAC ACACGGCGGG
1351 GTTTGA

```

WO 99/24578

PCT/IB98/01665

-137-

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```

1  MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
51 ADPHNLGWFF ILPPLAGMLV QPIVGHYS DR TWKPRLGRR LPYLLYGTLI
101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
5  151 DMVNEEQKGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
201 VVVAFYVGAA LLVITSAFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYQ
301 EAGNWyGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFESV
351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLENGS
10 401 ICMPQIVASL LSFVLFPLMG GLQATMFLVG GVVLLGAFS VFLIKETHGG
451 V*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N.*

15 *meningitidis*:

```

                                10      20      30
orfl6.pep                      GHYSDRTWKPR LXGRR LPYLLYGT LI AVIV
                                |||||
20 orfl6a                      IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPR LGRR LPYLLYGT LI AVIV
                                50      60      70      80      90      100

                                40      50      60      70      80      90
orfl6.pep                      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI
                                |||||
25 orfl6a                      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKGYAYGI
                                110     120     130     140     150     160

                                100     110     120     130     140     150
orfl6.pep                      QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQT VVVAFYVGAALLVITSAFTIFKVK
                                |||||
30 orfl6a                      QSFLANTGAVVAAILPFVFAYIGLANTA EKGVPQT VVVAFYVGAALLVITSAFTIFKVK
                                170     180     190     200     210     220

                                160     170     180
orfl6.pep                      EYXPETYARYHGIDVAANQEKANWIAL LKXA
                                || |||||
35 orfl6a                      EYNPETYARYHGIDVAANQEKANWIEL LKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI
                                230     240     250     260     270     280

40 orfl6a                      AENVWHTTDASSVGYQEAGNWyGVLA AVQSVAVICSFVLAKVPNKYHKAGYFGCLALGA
                                290     300     310     320     330     340

```

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

```

1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
51 AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
45 101 CCTTTACCCT GCAAAGCTCG CAGATGAGCC GCATCTCCA GACGCTCGGT
151 GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG
201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
251 CGCGTTTGGG CGGCCGCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
50 351 CTATGCGTCG CTGGCGGCTT TGTCGTTTCGG CGCGCTGATG ATTGCGCTGT
401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501 CTTAGCGAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
55 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
601 GTGGTCGTGG CGTTTTATGT GGGTGC GGCG TGTGCTGGTGA TTACCAGCGC
651 GTTACAGATT TTCAAAGTGA AGGAATACAA TCCGGAAACC TACGCCCGTT
701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
751 CTCTTGAAAA CCGCGCCTAA GGCGTTTTCG ACGGTTACTT TGGTGCAATT
801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
60 851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG

```

WO 99/24578

PCT/IB98/01665

-138-

5
1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
1101 CATCGCTTGG GCGGGCATTG TCACTTATCC GCTGACGATT GTGACCAACG
1151 CCTTGTCGGG CAAGCATATG GGCACCTACT TGGGCCTGTT TAACGGCTCT
1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGTCGTCC
1301 TGCTGCTGGG CGCGTTTTCG GTGTTCTGA TTAAAGAAAC ACACGGCGGG
1351 GTTTGA

This encodes a protein having amino acid sequence <SEQ ID 152>:

10
1 MSEYTPQTAQ QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
51 ADPHSLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI
101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
151 DMVNEEQKGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
201 VVVAFYVGAA LLVITSFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE
15
251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYYS AGAIAENVWH TTDASSVGYQ
301 EAGNWyGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNGS
401 ICMPOIVASL LSFVLFPM LG GLQATMFLVG GVVLLLGAFS VFLIKETHGG
451 V*

20 ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

		10	20	30	40	50	60
	orf16a.pep	MSEYTPQTAQ	QGLPALAKST	IWMLSFGFLG	VQTAFTLQSS	QMSRIFQTLG	ADPHSLGWFF
25	orf16-1	MSEYTPQTAQ	QGLPALAKST	IWMLSFGFLG	VQTAFTLQSS	QMSRIFQTLG	ADPHNLGWFF
		10	20	30	40	50	60
	orf16a.pep	70	80	90	100	110	120
		ILPPLAGMLV	QPIVGHYSR	TWKPRLGRR	LPYLLYGTLI	AVIVMILMPN	SGSFGFGYAS
30	orf16-1	ILPPLAGMLV	QPIVGHYSR	TWKPRLGRR	LPYLLYGTLI	AVIVMILMPN	SGSFGFGYAS
		70	80	90	100	110	120
	orf16a.pep	130	140	150	160	170	180
		LAALSFGALM	IALLDVSSNM	AMQPFKMMVG	DMVNEEQKGY	AYGIQSFLAN	TGAVVAAILP
35	orf16-1	LAALSFGALM	IALLDVSSNM	AMQPFKMMVG	DMVNEEQKGY	AYGIQSFLAN	TGAVVAAILP
		130	140	150	160	170	180
	orf16a.pep	190	200	210	220	230	240
		FVFAYIGLAN	TAEKGVVPQT	VVVAFYVGAA	LLVITSFTI	FKVKEYNPET	YARYHGIDVA
40	orf16-1	FVFAYIGLAN	TAEKGVVPQT	VVVAFYVGAA	LLVITSFTI	FKVKEYNPET	YARYHGIDVA
		190	200	210	220	230	240
	orf16a.pep	250	260	270	280	290	300
		ANQEKANWIE	LLKTAPKAFW	TVTLVQFFCW	FAFQYMWYYS	SAGAIAENVW	HTTDASSVGYQ
45	orf16-1	ANQEKANWIE	LLKTAPKAFW	TVTLVQFFCW	FAFQYMWYYS	SAGAIAENVW	HTTDASSVGYQ
		250	260	270	280	290	300
	orf16a.pep	310	320	330	340	350	360
		EAGNWyGVLA	AVQSVAVIC	SFVLAKVPNK	YHKAGYFGCL	ALGALGFFSV	FFIGNQYALV
50	orf16-1	EAGNWyGVLA	AVQSVAVIC	SFVLAKVPNK	YHKAGYFGCL	ALGALGFFSV	FFIGNQYALV
		310	320	330	340	350	360
	orf16a.pep	370	380	390	400	410	420
		LSYTLIGIAW	AGIITYPLTI	VTNALSGKHM	GTYLGLFNGS	ICMPQIVASL	LSFVLFPM LG
55	orf16-1	LSYTLIGIAW	AGIITYPLTI	VTNALSGKHM	GTYLGLFNGS	ICMPQIVASL	LSFVLFPM LG
		370	380	390	400	410	420
	orf16a.pep	430	440	450			
		GLQATMFLVG	GVVLLLGAFS	VFLIKETHGG	VX		
60	orf16-1	GLQATMFLVG	GVVLLLGAFS	VFLIKETHGG	VX		
		430	440	450			
65	orf16a.pep	430	440	450			
		GLQATMFLVG	GVVLLLGAFS	VFLIKETHGG	VX		
	orf16-1	GLQATMFLVG	GVVLLLGAFS	VFLIKETHGG	VX		
		430	440	450			

WO 99/24578

PCT/IB98/01665

-139-

Homology with a predicted ORF from *N.gonorrhoeae*

ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N. gonorrhoeae*:

5	orf16.pep	GHYSDRTWKPRLXGRRLPYLLYGTLIAVIV	30
	orf16ng	HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYSDRTWKPRLGGRRLPYLLYGTLIAVIV	131
10	orf16.pep	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKXYAYGI	90
	orf16ng	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKSYAYGI	191
15	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK	150
	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAALLIITSAFTISKVK	251
	orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
	orf16ng	EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWTVTPVQFFCWFAFRYMWTYSAGAI	311

20 The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
51	TACTTTTCAA	ATCAAAAAAA	AGGATTTACT	TTATGTCGGA	ATATACGCCT
101	CAAACAGCAA	AACAAGGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTTGGAT
151	GTTGAGCTTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAGCT
25	201	CGCAGATGAG	CCGCATTTT	CAAACGCTAG	GCGCAGACCC
	251	GGCTGGTTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG
	301	AGTGGCTACT	ACTCAGACCG	CACTTGGAAG	CCGCGCTTGG
	351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTCATC
	401	TGATGCCGAA	CTCGGCGAGC	TTCGGTTTCG	GCTATGCGTC
30	451	TTGTGCTTCG	GCGCGCTGAT	GATTGCGCTG	TTGGACGTGT
	501	GGCGATGCAG	CCGTTTAAAG	TGATGGTCGG	CGATATGGTC
	551	AGAAAAGCTA	CGCCTACGGG	ATTCAAAGTT	TCTTAGCGAA
	601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTTT	GCGTATATCG
	651	CACTGCCGAG	AAAGGCGTTG	TGCCACAAAC	CGTGGTCGTA
35	701	TGGGTGCGGC	GTTACTGATT	ATTACCAGTG	CGTTCACAA
	751	AAAGAATACG	ACCCGGAAAC	CTACGCCCGT	TACCACGGCA
	801	CGCGAATCAG	GAAAAAGCCA	ACTGGTTCGA	ACTCTTAAAA
	851	AAGTGTTTTG	GACGGTTACT	CCGGTACAGT	TTTTCTGCTG
	901	CGGTATATGT	GGAATTACTC	GGCAGGCGCG	ATTGCAGAAA
40	951	CACTACCGAT	GCGTCTCCG	TAGGCCATCA	GGAGGCGGGC
	1001	GCGTTTTGGC	GGCGGTGTAG		

This encodes a protein having amino acid sequence <SEQ ID 154>:

1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLLYVG	IYASNSKTRF	ARAGKKHDL
51	VELRLSRRSD	GLYPAKLADE	PHFSNARRRP	AQFGLVFHPA	AAGGDAGSAD
45	101	SGYSDRTWK	PRLGGRRLPY	LLYGTIAVI	VMILMPNSGS
	151	LSFGALMIAL	LDVSSNMAMQ	PFKMMVGDV	NEEQKSYAYG
	201	VVAAILPFVF	AYIGLANTAE	KGVPQTVV	AFYVGAALLI
	251	KEYDPETYAR	YHGIDVAANQ	EKANWFELLK	TAPKVFWTVT
	301	RYMWTYSAGA	IAENVHHTTD	ASSVGHQEAG	NRYGVLAHV*

50 ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

		30	40	50	60	70	80
	orf16-1.pep	MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSDRT					
	orf16ng	DVELRLSRRSDGLYPAKLADEPHFSNARRRP AQFGLVF-HPAAAGGDAGSADSGYSDRT					
55		50	60	70	80	90	100
	orf16-1.pep	90	100	110	120	130	140
	orf16ng	110	120	130	140	150	160

WO 99/24578

PCT/IB98/01665

-140-

		150	160	170	180	190	200
5	orf16-1.pep	MQPFKMMVGDMVNEEQKGYAYGIQSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTV					
	orf16ng	MQPFKMMVGDMVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTV					
		170	180	190	200	210	220
10	orf16-1.pep	210	220	230	240	250	260
	orf16ng	VVAFYVGAALLVITSFTIFKVKEYDPETYARYHGIDVAANQEKANWIELLKTAPKAFWT					
		230	240	250	260	270	280
15	orf16-1.pep	270	280	290	300	310	320
	orf16ng	VTLVQFFCWFAFQYMWYTSAGAI AENVWHTTDASSVGYQEAGNRYGVLA AVQSVA AVICS					
		290	300	310	320	330	340

- 20 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

25	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGCATA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	NAAACACGTT	GNCAAAGACC	AAATCCGNGN	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AA.NTGACGG
30	251	GNATTTTGAN	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TNAGGATACC
	301	CCGAGCTATG	C.TGCCACCA	AGCCCTGCCG	GTCAAACCTCG	GATCGNCTGG
	351	CAGCCAGAAT...				

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

35	1	MLFRKTTAAV	LAHTLMLNGC	TLMLWGMNNP	VSETITRKHV	XKDQIRXFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WVFNVPEDSA	XXTGILXAGL	DKPFQIVXDT
	101	PSYXCHQALP	VKLGSXGSQN...			

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

40	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGA	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TGAGGATACC
	301	CCGAGCTATG	CTCGCCACCA	AGCCCTGCCG	GTCAAACCTCG	AATCGCCTGG
45	351	CAGCCAGAAT	TTCAGTACCG	AAGGCCTTTG	CCTGCGCTAC	GATACCGACA
	401	AGCCTGCCGA	CATCGCCAAG	CTGAAACAGC	TCGGGTTTGA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAAATA
	501	CTACGCCACA	CCGCAAAAAC	TGAACGCCGA	TTACCATTTT	GAGCAAAGTG
50	551	TGCCTGCCGA	TATTTATTAC	ACGGTTACTG	AAGAACATAC	CGACAAATCC
	601	AAGCTGTTTG	CAAATATCTT	ATATACGCCC	CCCTTTTGA	TACTGGATGC
	651	GGCGGGCGCG	GTACTGGCCT	TGCCTGCGGC	GGCTCTGGGT	GCGGTCGTGG
	701	ATGCCGCCCG	CAATGA			

This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

55	1	MLFRKTTAAV	LAATLMLNGC	TLMLWGMNNP	VSETITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WVFNVPEDSA	KLTGILKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKLESPGSQN	FSTEGLCRLY	DTDKPADIAC	LKQLGFEAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIYY	TVTEEHTDKS

		10	20	30	40	50	60
45	orf28a.pep	MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK					
	orf28-1	MLFRKTTAAVLAATLMLNGCTLMLWGMNNPVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
50	orf28a.pep	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKQFQMVEPNRFA-YQALPVKLESPASQN					
	orf28-1	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPG SQN					
		70	80	90	100	110	119
55	orf28a.pep	FSTEGLC LRYDTRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF					
	orf28-1	FSTEGLC LRYDTRPADIAKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		120	130	140	150	160	170
		130	140	150	160	170	180

WO 99/24578

PCT/IB98/01665

-142-

		180	190	200	210	220	230
	orf28a.pep	EQSVPADIIYYTVTKKHTDKSKLFENIAYTPPTLILDAVGAVLALPVAALIAATNSSDKX					
		: : : : : : : :					
5	orf28-1	EQSVPADIIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.*10 *gonorrhoeae*:

	orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK	60
		: : : : : : :	
	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKDQIRAFGVVAEDNAQLEK	60
15	orf28.pep	GSLVMMGGKYWFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXGSON	120
		: : : : : : :	
	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON	120

The complete length ORF28ng nucleotide sequence <SEQ ID 161> is

	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATACT
20	51	GAACGGCTGT	ACGATGATGT	TGCGGGGGAT	GAACAACCCG	GTCAGCCAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGGAAATAC	TGGTTCGCCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
	251	GCCTTTTGAA	GGCCGGGTG	GACAAGCCCT	TCCAAATAGT	TGAGGATACC
25	301	CCGAGCTATG	CCCGCCACCA	AGCCCTGCCG	GTCAAATTCG	AAGCGCCCGG
	351	CAGCCAGAAT	TTCAGTACCG	GAGGTCTTTG	CCTGCGCTAT	GATACCGGCA
	401	GACCTGACGA	CATCGCCAAG	CTGAAACAGC	TTGAGTTTAA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTA	CACGCGCTGC	GTATCCGCCA	AAGGCAAATA
	501	CTACGCCACG	CCGCAAAAAC	TGAACGCCGA	TTATCATTTT	GAGCAAAGTG
30	551	TGCCCGCCGA	TATTTATTAT	ACGGTTACTG	AAAAACATAC	CGACAAATCC
	601	AAGCTGTTTG	GAAATATCTT	ATATACGCC	CCCTTGTTGA	TATTGGATGC
	651	GGCGGCCGCG	GTGCTGGTCT	TGCCTATGGC	TCTGATTGCA	GCCGCGAATT
	701	CCTCAGACAA	ATGA			

This encodes a protein having amino acid sequence <SEQ ID 162>:

35	1	MLFRKTTAAV	LAATLILNGC	TMMLRGMNPN	VSQTITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFAVNPEDSA	KLTGLLKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKFEAPGSON	FSTGGLCLRY	DTGRPDDIAK	LKQLEFKAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIIY	TVTEKHTDKS
	201	KLFGNILYTP	PLLILDAAAA	VLVLPMALIA	AANSSDK*	

40 ORF28ng and ORF28-1 share 90.0% identity in 231 aa overlap:

		10	20	30	40	50	60
	orf28-1.pep	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
		: : : : : :					
45	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKDQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
	orf28-1.pep	70	80	90	100	110	120
		GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSON					
		: : : : : :					
50	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON					
		70	80	90	100	110	120
	orf28-1.pep	130	140	150	160	170	180
		FSTEGCLCLRYDTDKPADIAPKQKLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
55		: : : : : :					
	orf28ng	FSTGGLCLRYDTGRPDDIAKQKLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		130	140	150	160	170	180
	orf28-1.pep	190	200	210	220	230	239
60		EQSVPADIIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
		: : : : : :					
	orf28ng	EQSVPADIIYYTVTEKHTDKSKLFGNILYTPPLLILDAAAAVLVLPMALIAAANSSDKX					

190 200 210 220 230

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

```

15      1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
      51  TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
      101  CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTTCAG CGGCGGTGTA
      151  GACGGCGGTT TTACTGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
      201  TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCCG
      251  GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
20      301  ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
      351  AGAAAATGCC GGTGCCGCCT CTGGT..

```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```

25      1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHHDS KSTSDFSGGV
      51  DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYYVKGTSTK
      101  TKTSIVPQAP FSDRWLEENA GAASG..

```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

```

30      1  ATGAATTTGC CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC
      51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
      101  GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
      151  TTTGGTAATG CTCGCGGCAG TGTTAAAAAG CGGGTTTACG CCGTCCAGAC
      201  ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
      251  CAGGGTTTGA AGGTGTTATC GGTTATGAAA CCCATTTTTC AGGGCACGGA
      301  CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
      351  TTTTCAGCGC GGTGTAGACG GCGGTTTTAC TGTTTACCAA CTTTCATCGAA
35      401  CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
      451  GATTATCCGC CCCCCGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
      501  AGGAACTTCA ACAAAAACAA AGACTAATAT TGTCCCTCAA GCCCCATTTT
      551  CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
      601  CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
40      651  TTGGTGGGCT AACCGTATGG ATGATGTTTC CGGCATCGTC CAAGGTGCGG
      701  TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
      751  GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
      801  AGGTATTAAT GATTTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG
      851  CGAGCCTATT ACAGGACAGT GCTTTTGCGG TAAAAGACGG TATCAACTCT
45      901  GCCAAACAAT GGGCTGATGC CCATCCAAAT ATAACAGCTA CTGCCCAAAC
      951  TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTTGGAGA GGTAAGAAAG
      1001  TAGAACTTAA CCCGACTAAA TGGGATTGGG TTAAAAATAC CGGTTATAAA
      1051  AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
      1101  GAATAAACCT ATTAATCTT TACCAAACAG TGCCGCTGAA AAAAGAAAAC
50      1151  AAAATTTTGA GAAGTTTAAT AGTAACTGGA GTTCAGCAAG TTTTGATTCA

```

WO 99/24578

PCT/IB98/01665

-144-

5
1201 GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA
1251 AGTTAAAACT CGATACACTA GTTTAGATGG AAAAATTACA ATTATAAAAG
1301 ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCACG AAAACAGTAT
1351 CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAATTAC AAGGTAAGCA
1401 AGCAAAAGAT TATTTACAAC AACAACTCA TATCAGGAAC TTAGACAAAT
1451 GA

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

10
15
1 MNLPIQKFM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPPGKYHL
51 FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG
101 HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151 DYPPPGGARD IYSYVKGTS TKTCTNIVPQ APFSDRWLKE NAGAASGFFS
201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT
251 DSAVSPVTD TAAQQTLOGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS
301 AKQWADAHN ITATAQTALS AAEAAGTVWR GKKVELNPTK WDWVKNITGYK
351 KPAARHMQTL DGEMAGGNKP IKSLPNSAAE KRKQNFEEK SNWSSASFDS
401 VHKTLPNAP GILSPDKVKT RYTSLDGKIT IIKDNENNYF RIHDNSRKQY
451 LDSNGNAVKT GNLOGKQAKD YLQQQTHIRN LDK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis*:

				10	20	30
	orf29.pep			VSPVLPIT	HERTGFEGVIGYETHFSGHGHE	
				: :	: :	: :
25	orf29a	EPGGKYHLFGNARGSVKNRVYAVQTFDATAVGPILPIT	HERTGFEGVIGYETHFSGHGHE			
		50	60	70	80	90
		40	50	60	70	80
30	orf29.pep	VHSPFDHHD	SKSTSDFSGGVDGGFTVYQLHRTWSEIHP	PEDEYDGPQAAXYPPPGGARDIY		
		: :	: :	: :	: :	: :
	orf29a	VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHP	EDGYDGPQGS	DYPPPGGARDIY		
		110	120	130	140	150
35	orf29.pep		100	110	120	
			SYVKGTS	TKTKTSIVPQAPFSDRWLEENAGAASG		
			: :	: :	: :	: :
	orf29a		XXVKGTS	TKTKSNIVPRAPFSDRWLKENAGAASGFFS	RADEAGKLIWESDPNKNWWANR	
		170	180	190	200	210
40	orf29a		MDDIRGIVQGAVNPFLMG	FQGVGIGAIT	DSAVSPVTD	TAAQQTLOGXNHLGXLSPEAQLA
		230	240	250	260	270

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

45
50
55
60
1 ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC
51 GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
151 TTTGGTAATG CTCGCGGCAG TGTTAAAAAT CGGGTTTACG CCGTCCAAAC
201 ATTTGATGCA ACTGCGGTCG GCGCCATACT GCCTATTACA CACGAACGGA
251 CAGGATTTGA AGGCATTATC GGTTATGAAA CCCATTTTTC AGGACATGGA
301 CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA
351 TTTGAGCGGC GCGGTAGACG GTGGTTTAC CGTTTACCAA CTTTCATCGGA
401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
451 GATTATCCGC CCGCGGAGG AGCAAGGGAT ATATACANNT ANTATGTCAA
501 AGGAACTTCA ACAAACAAA AGAGTAATAT TGTTCCCGCA GCCCATTTT
551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
651 TTGGTGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG
701 TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
801 AGGTATNAAT CATTAGGAA ANTAAAGTCC CGAAGCACAA CTTGCGGCTG
851 CAACCGCATT ACAAGACAGT GCTTTTGCAG TAAAAGACGG TATCAATTCC
901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAAGTCAA CAGCCCAAAC

WO 99/24578

PCT/IB98/01665

-146-

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N. gonorrhoeae*:

	orf29.pep	VSPVLPITHERTGFEVIGYETHFSGHGHE	30
5	orf29ng	EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEVIGYETHFSGHGHE	102
	orf29.pep	VHSPFDHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDGYDGPQAAXYPPPGGARDIY	90
10	orf29ng	VHSPFDNHDSDKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGGYPPPGGARDIY	162
	orf29.pep	SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG	125
	orf29ng	SYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADKLIWENDPDKNWRANR	222

15 The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

	1	MNLPIQKFMM	LFAAAISLLQ	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
	51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEVIG	GYETHFSGHG
20	101	HEVHSPFDNH	DSKSTSDFSG	GVDGGFTVYQ	LHRTGSEIHP	EDGYDGPQGG
	151	GYPPPGGARD	IYSYHIKGT	TKTKINTVPQ	APFSDRWLKE	NAGAASGFLS
	201	RADEAGKLIW	ENDPDKNRA	NRMDDIRGIV	QGA VNPFLT	FGGLGVGAIT
	251	DSAVSPVTYA	AARKTLQGIH	NLGNLSPEAQ	LAAATALQDS	AFAVKDSINS
	301	ARQWADAHPN	ITATAQTALA	VTEAATTVWG	GKKVELNPAK	WDWVKNTGYK
	351	KPAARHMOTV	DGEMAGGNKP	LESKNTVTN	NFFENTGYTE	KVLRQASNGD
25	401	YHGFPQSVDA	FSENGTVIQI	VGGDNIVRHK	LYIPGSYKKG	DGNFEYIREA
	451	DGKINHRLFV	PNQQLPEK*			

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

	1	atgAATTTGC	CTATTCAAAA	ATTCATGATG	ctgttggcAg	cggcaatatc
	51	gatgctGCat	ATCCCCATTA	GTCATGCGAA	CGGTTTGGAT	GCCCCGTTTGC
30	101	GCGATGATAT	GCAGGCAAAA	CACTACGAAC	CGGGTGGCAA	ATACCATCTG
	151	TTTGGTAATG	CTCGCGGCAG	TGTTAAAAAT	CGGGTTTGCG	CCGTCCAAAC
	201	ATTTGATGCA	ACTGCGGTCG	GCCCCATACT	GCCTATTACA	CACGAACGGA
	251	CAGGATTTGA	AGGTGTTATC	GGCTATGAAA	CCCATTTTTC	AGGACACGGA
	301	CACGAAGTAC	ACAGTCCGTT	CGATAATCAT	GATTCAAAAA	GCACTTCTGA
35	351	TTTCAGCGGC	GGCGTAGACG	GCGGTTTAC	CGTTTACCAA	CTTCATCGGA
	401	CAGGTCGGA	AATACATCCC	GCAGACGGAT	ATGACGGGCC	TCAAGGCGGC
	451	GGTTATCCGG	AACCACAAGG	GGCAAGGGAT	ATATACAGCT	ACCATATCAA
	501	AGGAACTTCA	ACCAAAACAA	AGATAAACAC	TGTTCCGCAA	GCCCCTTTTT
	551	CAGACCGCTG	GCTAAAAGAA	AATGCCGGTG	CCGCTTCCGG	TTTTCTCAGC
40	601	CGTGCGGATG	AAGCAGGAAA	ACTGATATGG	GAAAACGACC	CCGATAAAAA
	651	TTGGCGGGCT	AACCGTATGG	ATGATATTCG	CGGCATCGTC	CAAGGTGCGG
	701	TTAATCCTTT	TTTAACGGGT	TTCAAGGGG	TAGGGATTGG	GGCAATTACA
	751	GACAGTGCGG	TAAGCCCGGT	CACAGATACA	GCCGCTCAGC	AGACTCTACA
	801	AGGTATTAAT	GATTTAGGAA	ATTTAAGTCC	GGAAGCACAA	CTTGCCGCCG
45	851	CGAGCCTATT	ACAGGACAGT	GCCTTTGCGG	TAAAAGACGG	CATCAATTCC
	901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAACAGCAA	CAGCCCAAAC
	951	TGCCCTTGCC	GTCAGCAGAG	CCGCAGGTAC	GGTTTGCGC	GGTAAAAAAG
	1001	TAGAACTTAA	CCCGACCAAA	TGGGATTGGG	TTAAAAATAC	CGGCTATAAA
	1051	AAACCTGCTG	CCCGCCATAT	GCAGACTGTA	GATGGGGAGA	TGGCAGGGGG
50	1101	GAATAGACCG	CCTAAATCTA	TAACGTCGGA	AGGAAAAGCT	AATGCTGCAA
	1151	CCTATCCTAA	GTTGGTTAAT	CAGCTAAATG	AGCAAACTT	AAATAACATT
	1201	GCGGCTCAAG	ATCCAAGATT	GAGTCTAGCT	ATTCATGAGG	GTAAAAAATA
	1251	TTTTCCAATA	GGAAGTCAAA	CTTATGAAGA	GGCAGATAGA	CTAGGTAAAA
	1301	TTTGGGTTGG	TGAGGGTGCA	AGACAAACTA	GTGGAGGCGG	ATGGTTAAGT
55	1351	AGAGATGGCA	CTCGACAATA	TCGGCCACCA	ACAGAAAAAA	AATCACAATT
	1401	TGCAACTACA	GGTATTCAAG	CAAATTTTGA	AACTTATACT	ATTGATTCAA
	1451	ATGAAAAAAG	AAATAAAATT	AAAAATGGAC	ATTTAAATAT	TAGGTAA

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

	1	MNLPIQKFMM	LLAAISMHLH	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
60	51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEVIG	GYETHFSGHG

WO 99/24578

PCT/IB98/01665

-148-

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAG...

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

5 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

10 1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTC
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 151 ATGAAGGAGA CAGAGGGGGC GTTCTTCCA TTGGCTATCT TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
 15 401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

20 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKETEGAFLEP LAILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
 101 PGGVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
 151 DNTGKTLPGQ GIGRHPWES KSTDRSWKNR F*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

30 orf30.pep 10 20 30 40
 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
 |||||
 orf30a MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
 10 20 30 40 50 60
 orf30a LXILGGAAIGMWTQHGFSA TTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI
 70 80 90 100 110 120

35 The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

40 1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTC
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 151 ATGAAGGANA CAGNGGGGGC GTTCTTCCA TTGGNTATCT TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 301 CCTGGTGN TG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
 401 GAACAGGTCA TCCTATTGGN AAATTTCCCC ATTATCATCG TCGAGTTACG
 45 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

This encodes a protein having amino acid sequence <SEQ ID 178>:

50 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKXTXGAFLP LXILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
 101 PGXVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
 151 DNTGKTLPGQ GIGRHPWES KSTDRSWKNR F*

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

orf30a.pep MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60

	orf30-1	 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60
5	orf30a.pep	LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI	120
	orf30-1	LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI	120
10	orf30a.pep	KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
	orf30-1	KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
	orf30a.pep	FX	
	orf30-1	FX	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae*:

20	orf30.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ	42
		:	
	orf30ng	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

25	1	ATGAAAAAAC	AAATCACCGC	AGCCGTAATG	ATGCTGTCTA	TGATCGCCCC
	51	CGCAATTGGCA	AACGGATTGG	ACAATCAGGC	ATTTGAAGAC	CAAGTGTTCC
	101	ACACGCGGGC	AGATGCGCCG	ATGCAGTTGG	CGGAGCTTTC	TCAGAAGGAG
	151	ATGAAGGAGA	CTGAAGGGGC	TTTTCTTCCA	TTGGCTATCT	TGGGTGGTGC
	201	TGCCATTGGT	ATGTGGACAC	AGCATGGTTT	TAGTTATGCA	ACGACAGGCA
30	251	GACCAGCTTC	TGTTAGAGAT	GTTGCTGGCG	GATTAGGCGC	AATTCCTGGT
	301	GATGTAGGTG	CTGCAGGAAA	GGTTGTTTCC	TTTGCTAAAT	ATGGACGTGA
	351	GATTAAATC	GGCAATAATA	TGCGGATAGC	CCCTTTCGGT	AATAGAACAG
	401	GTCATCCTAT	TGGAATTTT	CCCCATTATC	ATCGTCGAGT	TACGGATAAT
	451	ACGGGCAAGA	CTTTGCCTGG	ACAGGGAATT	GGTCGTCATC	GCCCTTGGGA
	501	ATCAAATCT	ACGGACAGAT	CATGGAAAAA	CCGCTTCTAA	

This encodes a protein having amino acid sequence <SEQ ID 180>:

35 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKETEGAFLP LAILGGAAIG MWTQHGFSYA TTGRPASVRD VAGGLGAIPG
 101 DVGGAAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN
 151 TGKTLPGOGI GRHRPWESKS TDRSWKNRF*

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

40		10	20	30	40	50	60
	orf30ng.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP					
	orf30-1	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP					
		10	20	30	40	50	60
45		70	80	90	100	110	
	orf30ng.pep	LAILGGAAIGMWTQHGFSYATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI					
	orf30-1	LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI					
		70	80	90	100	110	120
50		120	130	140	150	160	170
	orf30ng.pep	KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR					
55	orf30-1	KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR					
		130	140	150	160	170	180
	orf30ng.pep	180	FX				
60	orf30-1	FX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 22

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

10

```

      1  ATGAATAAAA  CTCTCTATCG  TGTAATTTTC  AACCGCAAAC  GTGGGGCTGT
     51  GrTAGCCGTT  GCTGAAACTA  CCAAGCGCGA  AGGTAAAAGC  TGTGCCGATA
    101  GTGATTCAGG  CAGCGCTCAT  GTGAAATCTG  TTCCTTTTGG  TACTACTCAT
    151  GCACCTGTTT  GTg.CGTTaC  AAATATCTTT  TCTTTTTTCTT  TATTGGGCTT
    201  TTCTTTATGT  TTGGCTGTAG  GtacGGyCAA  TATTGCTTTT  GCTGATGGCA
    251  TT..

```

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

1 MNKTLYRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVPEFGTH
51 APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..

- 15 Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

20

1	ATGAATAAAA	CTCTCTATCG	TGTAATTTTC	AACCGCAAAC	GTGGGGCTGT
51	GGTGACCGTT	GCTGAAACTA	CCAAGCGCGA	AGGTAAAAGC	TGTGCCGATA
101	GTGATTCAGG	CAGCGCTCAT	GTGAAATCTG	TTCCTTTTGG	TACTACTCAT
151	GCACCTGTTT	GTCGTTCAAA	TATCTTTTCT	TTTTCTTTAT	TGGGCTTTTC
201	TTTATGTTTG	GCTGTAGGTA	CGGCCAATAT	TGCTTTTGCT	GATGGCATT..

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

1 MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPFGTTH
51 APVCRSNIFS FSLGFSLCL AVGTANIAFA DGI..

Computer analysis of this amino acid sequence gave the following results:

- 25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N. gonorrhoeae*:

	orf31.pep	MNKTLYRVIFNRKRGAVXAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF	60
		:	
30	orf31.ng	MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAF	54
	orf31.pep	SFSLLGFSLCLAVGTXNIAFADGI	84
		:	
	orf31.ng	CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSV	114

- 35 The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

40	1	ATGAACAAA	CCCTCTATCG	TGTGATTTTC	AACCGCAAAC	GCGGTGCTGT
	51	GGTAGCTGTT	GCCGAAACCA	CCAAGCGCGA	AGGTAAGC	TGTGCCGATA
	101	GTGGTTCGGG	CAGCGTTTAT	GTGAAATCCG	TTTCTTTCAT	TCCTACTCAT
	151	TCCAAAGCCT	TTTGTTTTTC	TGCATTAGGC	TTTTCTTTAT	GTTTGGCTTT
	201	GGGTACGGTC	AATATTGCTT	TTGCTGACGG	CATTATTACT	GATAAAGCTG
	251	CTCCTAAAAC	CCAACAAGCC	ACGATTCTGC	AAACAGGTaa	cGGCATACCG
	301	CAAGTCAATA	TTCAAACCCC	TACTTCGGCA	GGGGTTTCTG	TTAATCAATA
45	351	TGCCCAGTTT	GATGTGGGTA	ATCGCGGGGC	GATTTTAAAC	AACAGTCGCA
	401	GCAACACCCA	AACACAGCTA	GGCGGTTGGA	TTCAAGGCAA	TCCTTGGTTG
	451	ACAAGGGGCG	AAGCACGTGT	GGTTGTAAAC	CAAATCAACA	GCAGCCATCC
	501	TTCACAACTG	AATGGCTATA	TTGAAGTGGG	TGGACGACGT	GCAGAAGTCG
	551	TTATTGCCAA	TCCGGCAGGG	ATTGCAGTCA	ATGGTGGTGG	TTTTATCAAT
	601	GCTTCCCGTG	CCACTTTGAC	GACAGGCCAA	CCGCAATATC	AAGCAGGAGA
	651	CTTTAGCGGC	TTTAAGATAA	GGCAAGGCAA	TGCTGTAATC	GCCGGACACG

50

1	ATGAATACTC	CTCCTTTTGT	CTGTTGGATT	TTTTGCAAGG	TCATCGACAA
51	TTTCGGCGAC	ATCGGCGTTT	CGTGGCGGCT	CGCCCGTGTT	TTGCACCGCG
101	AACTCGGTTG	GCAGGTGCAT	TTGTGGACGG	ACGATGTGTC	CGCCTTGCCT

WO 99/24578

PCT/IB98/01665

-152-

5
10
15
20

```

151 GCGCTTTGCC CTGATTTGCC CGATGTTCCC TCGGTTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCCGA AAATGTGCTG
301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
351 CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
401 GTGTTCAAAA ATATTTTGGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
501 CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
651 CAGCCTCAAA CAAAGCGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TGC GCGCCCA GCTTGC GGGC AAACCCTTCT
851 TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGTTTCTAC ACGCCCGAAA CCGTGTGCGC
951 ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAACCCTGC AACAACATCA AAACGGCTGG
1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
1101 TCCTGAAAAA CTCGCTGCCT TTGTTTCAA GCATCAAAAA ATACGCTAG

```

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

25

```

1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
151 LIRERDYCEA VRFDTEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
201 QAGSPMTLLL AGTQIIDSLE QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATQRLCW QTLOQHONGW
351 RQGAEDWSRY LFGQPSAPEK LAAFVSKHQK IR*w

```

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N. meningitidis*:

35

```

          10      20      30      40      50      60
orf32.pep MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
          |||||
orf32a    MNTPPFSAGXFCCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
          10      20      30      40      50      60

40
          70      80
orf32.pep CVHQDIHVRTWHSDAADIDTA
          |||||
orf32a    CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
          70      80      90      100     110     120

```

45 The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

50
55
60

```

1 ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
51 TTTTCGGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGTT TTGCACCGCG
101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151 GCGCTTTGCC CTGATTTGCC CGATGTTTNC TCGGTTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCCGA AAATGTGCTG
301 CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
351 CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
401 GTGTTCAAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
451 CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
501 CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNNGGGGCGC ANATTATCGA
651 CAGCCTCAAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT

```

WO 99/24578

PCT/IB98/01665

-153-

5
801 CCGCGGCGAA GACAGTTTCG TGC GCGCCCA GCTTGCGGGC AAACCCTTCT
851 TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
951 ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAATCCTGC AACACATCA AAACGGCTGG
1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC
1101 ATCCGAAAAA CTCGCCGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG

This encodes a protein having amino acid sequence <SEQ ID 192>:

10
1 1 MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51 ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL
101 HIIRRHKPLW LXWEYLSAEX SNERLHXMP S PQESVXKXFW FMGFSEXSGG
151 LIRERDYCEA VRFD SGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR
201 QAGSPLTLL AGAXIIDSLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
15 301 AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLECW QILQHQHNGW
351 RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

		10	20	30	40	50	60
20	orf32-1.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP					
	orf32a	MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX					
		10	20	30	40	50	60
25	orf32-1.pep	70	80	90	100	110	120
		CVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAEE					
	orf32a	CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX					
		70	80	90	100	110	120
30	orf32-1.pep	130	140	150	160	170	180
		SNERLHLMPS PQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS					
	orf32a	SNERLHXMPSPQESVXKXFWFMGFSEXSGGLIRERDYCEAVRFD SGALRKRLMLPEKNXP					
		130	140	150	160	170	180
35	orf32-1.pep	190	200	210	220	230	240
		EWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDVFQTA					
	orf32a	EWLLFGYRSDVWAKWLEMWRQAGSPLTLLLAGAXIIDSLKQNGVIPQDALQNDGDVFQTA					
40		190	200	210	220	230	240
45	orf32-1.pep	250	260	270	280	290	300
		SVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH					
	orf32a	SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH					
		250	260	270	280	290	300
50	orf32-1.pep	310	320	330	340	350	360
		AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLECWQTLQHQHNGWRQGAEDWSRY					
	orf32a	AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATQRLECWQILQHQHNGWRQGAEDWSRY					
		310	320	330	340	350	360
55	orf32-1.pep	370	380				
		LFGQPSAPEKLAAFVSKHQKIRX					
	orf32a	LFGQPSASEKLAAFVSKHQKIRX					
		370	380				

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N.gonorrhoeae*:

orf32.pep	MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP	57

WO 99/24578

PCT/IB98/01665

-154-

orf32ng MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP 60
 orf32.pep DVPCVHQDIHVRTWHSDAADIDTA 81
 orf32ng DVPFVHQDIHVRTWHSDAADIDTAPVDPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

1 MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDDVS
 51 ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVDAV IETFACDLPE
 101 NVLNIIRRHK PLWLNWEYLS AEESNERLHL MPSPQEGVQK YFWFMGFSEK
 151 SGGLIRERDY REAVRFDTEA LRRRLVLPEK NAPEWLLFGY RGDVWAKWLD
 201 MWQQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
 251 PFVPQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFWHI YPDENVHLD
 301 KLHAFWDKAY GFYTPETASV HRLSDDLNG GEALSATQRL ECGVL*

Further sequencing revealed the following DNA sequence <SEQ ID 195>:

1 ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTTTGCA AGGTCATCGA
 51 CAATTCGGC GACATCGGCG TTTCGTGGCG GCTCGCCCGT GTTTTGCACC
 101 GCGAACTCGG TTGGCAGGTG CATTTGTGGA CGGACGACGT GTCCGCCTTG
 151 CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT
 201 TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
 251 TTCCCGATGC CGTTATCGAA ACTTTTGCTT GCGACCTGCC CGAAAATGTG
 301 CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGGAATATTT
 351 GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGCAGG
 401 AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
 451 GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATAACGA
 501 AGCCCTGCGC CGGCGGCTGG TGCTGCCC GA AAAAACGCC CCCGAATGGC
 551 TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
 601 CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
 651 CGACAGCCTC AAACAAAGCG GCGTTATTCC GCAAAACGCC CTGCAAAAtg
 701 aaggcgtGTG CTTTCagacG gcatccgTcC gccttGTCAA AAtcCCGTTC
 751 GTGCcGCAAC AGGAcTTCGA CAAATTGCTG CAcctcgcCG ACTGCGCCGT
 801 GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCCCT
 851 TTTTTTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACAAACTC
 901 CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCC GAAACCGCATC
 951 GGTGCACCGC CTCCTTTCGG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
 1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC
 1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC
 1101 CGCATCCGAA AAACCTGCGC CTTTGTGTTT AAAGCATCAA AAAATACGCT
 1151 AG

This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

1 MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDDVSAL
 51 RALCPDLPDV PFVHQDIHVR TWHSDAADID TAPVDAVIE TFACDLPENV
 101 LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYF WFMGFSEKSG
 151 GLIRERDYRE AVRFDTEALR RRLVLPEKNA PEWLLFGYRG DVWAKWLDMW
 201 QQAGSLMTLL LAGAQIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
 251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDKL
 301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLC WQTLQHQHNG
 351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

50 orf32-1.pep 10 20 30 40 50 59
 MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDV
 orf32ng-1 10 20 30 40 50 60
 MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDV
 55 orf32-1.pep 60 70 80 90 100 110 119
 PCVHQDIHVRTWHSDAADIDTAPVDPDVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE
 orf32ng-1 60 70 80 90 100 110 120
 PFVHQDIHVRTWHSDAADIDTAPVDPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE
 60 120 130 140 150 160 170 179

WO 99/24578

PCT/IB98/01665

-155-

	orf32-1.pep	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNA					
	orf32ng-1	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYREAVRFDTEALRRRLVLPEKNA					
		130	140	150	160	170	180
5	orf32-1.pep	180	190	200	210	220	230
	orf32ng-1	180	190	200	210	220	230
10	orf32-1.pep	240	250	260	270	280	290
	orf32ng-1	240	250	260	270	280	290
15	orf32-1.pep	300	310	320	330	340	350
	orf32ng-1	300	310	320	330	340	350
20	orf32-1.pep	360	370	380			
	orf32ng-1	360	370	380			
25	orf32-1.pep	360	370	380			
	orf32ng-1	360	370	380			

30 On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 24

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

1 ..TTGTTCTGCG GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG
51 GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG
101 ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
151 AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTGCG TATTGTTGCT
45 201 GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA
251 ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTCGAAA
301 CTCGGTTTCC CTGTCCCCGA TCGCGGGTCG GTCATCGAAG GCCGTCTGAA
351 CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
401 TCGCCTGCTA NGGCATCCTG CCGCGCCTG..

50 This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

1 ..LFLRVKVGFR FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
51 SLWLCTLLGM LVSLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPK
101 LGFPVPDARS VIEGRLNGNI ADARAWSGLL VXSIACXGIL PRL..

WO 99/24578

PCT/IB98/01665

-156-

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```

1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
51 AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGGAGATG
5  151 ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TCGGTGCGGG
201 GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCATTT TTTACCGGTT
251 TTTCAGTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
301 GTTTTGGCGG GCGTGTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
10  401 CGACGTGGTT TCGGGGCAAA GACCCTGTAA ATCAGGCGGT GTTGCGGCTG
451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGGAAAA TAGGCGCAAC
501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 TGTTGCTGCT TTTGGTGC GG CAATATACGT TCAACTGGGA AAGCACGCTG
601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
15  651 GTCGAAACTC GGTTCCTCTG TCCCCGATGC GCGGGCGGTC ATCGAAGGCC
701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCGGG GCTGCTGGTC
751 GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTGCTGG CTTGGGTAGT
801 GTGTAAAATC CTTTTGAAAA CAAGCGAAAA CGGATTGGAT TTGGAAAAGC
851 CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
20  901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAAA TCATCTTGAA
951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGCGCAACTG CTTATCGGCG TCGCGCCCA AACTGTGCCG GACCGCGGCG
25  1151 TGTTGCGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCAGCTTT TGGCGGAACA GGGGCTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
1301 CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

```

30  1  MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
51  51  IDNRNMLRET LERVAGSFW LWVVAATFAF FTGFSVTYLL MDNQGLNFFL
101 101  VLAGVLGMNT LMLAVWLAML FLRVKVGRRF SSPATWFRGK DPNQAVLRL
151 151  YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201 201  LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRLNGNIA DARAWSGLLV
35  251  GSIACYGILP RLLAWVCKI LLKTSENGLD LEKPYQAVI RRWQNKITDA
301 301  DTRRETSAV SPKIILNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351 351  TNREQVALE TELKQKPAQL LIGVRAQTVP DRGVLROIIVR LSEAAQGGAV
401 401  VQLLAEQGLS DDLSEKLEHW RNALAECEGAA WLEPDRAAQE GRLKDQ*

```

Computer analysis of this amino acid sequence gave the following results:

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N. meningitidis*:

```

45  orf33.pep                                10      20      30
                                         LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR
                                         |||
orf33a      LMDNQGLNFFLVLAGVXGMNTLMLAVWLAMLFLRVKVGRRFFSSPATWFRGKDPVNQAVLR
          90      100      110      120      130      140

          40      50      60      70      80      90
orf33.pep  LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA
          |||
orf33a      LYADEWRXPSVRWKIGATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLGDSSSVRL
          150     160     170     180     190     200

          100     110     120     130     140
orf33.pep  VEMLAWLPSKLGFVPDARSVIEGRLNGNIA DARAWSGLLVXSIACXGILPRL
          |||
orf33a      VEMLAWLPAKLGFPVPDARAVIEGRLNGNIA DARAWSGLLVGSIACYGILPRLLA VAVCK
          210     220     230     240     250     260

60  orf33a      ILXXTSENGLDLEKXXXXXIRRWQNKITDADTRRETSAVSPKIVLNDAPKWAVMLETE
          270     280     290     300     310     320

```


WO 99/24578

PCT/IB98/01665

-158-

	orf33-1	DARAWSGLLVGS	250	260	270	280	290	300
			310	320	330	340	350	360
5	orf33a.pep	DTRRET	SAVSPKIVLNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGV	AANREQVAAL				
	orf33-1	DTRRET	SAVSPKII	LNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGV	ATNREQVAAL			
10	orf33a.pep	TELKQKPAQL	LIGVRAQTV	PDRGVLRQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW				
	orf33-1	TELKQKPAQL	LIGVRAQTV	PDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW				
15	orf33a.pep	RNALTECGAAWLEP	DRAAQEGRLKTNDRTX					
	orf33-1	RNALAECGA	AWLEPDR	AAQEGRLKDQX				
20			430	440				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N. gonorrhoeae*:

25	orf33.pep	LFLRVKVG	RFFSSPATWFRXKDPVNQAVLR	30
	orf33ng	LMDNQGLNFF	LVLAVWLATLFLRVKVG	100
30	orf33.pep	LYXDEWRXTSV	RWKIXATSHSLWLCTLLGMLVSVLLLLLV	90
	orf33ng	LYADQWRQPSV	RWKIGATAHSLWLCTLLGMLVSVLLLLLV	160
	orf33.pep	VEMLAWLPSK	LGFPVPDARSVIEGRLNGNIADARAWSGLLVXSIACXGILPRL	143
35	orf33ng	VEMLAWLPSK	LGFPVPDARAVIEGRLNGNIADARAWSGLLVGSIVCYGILPRL	220

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

	1	MIDRDRMLRD	TLERV	RAGSF	WLWVV	VASMM	FTAGF	SGTYL	LMDNQGLNFF
	51	LVLAVGLGMN	TLMLAVWLAT	LFLRVKVG	RFFSSPATWFRG	KGPVNQAVLR			
40	101	LYADQWRQPS	VRWKIGATAH	SLWLCTLLGM	LVSLLLLLV	RQYTFNWEST			
	151	LLSNAASVRA	VEMLAWLPSK	LGFPVPDARA	VIEGRLNGNI	ADARAWSGLL			
	201	VGSIVCYGIL	PRLAWVVCK	ILLKTS	ENGLDLEK	TPYYQAV	IRRWQNKITD		
	251	ADTRRET	VSAVSPKIVLND	APKWAVMLETE	WQDGQWFEGR	LAQEWLDKGV			
	301	AANREQVAAL	ETELKQKPAQ	LLIGVRAQTV	PDRGVLRQIV	RLSEAAQGGA			
45	351	VVQLLAEQGL	SDDLSEKLEH	WRNALTECGA	AWLEPDRVAQ	EGRLKDQ*			

Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

	1	ATGTTGa	atC	CATCCCga	AA	ACTGgtt	gag	ctGgTCC	GtA	Ttttga	ataa
	51	aggggg	gtTTT	attttc	agcg	gcgatc	ctgt	gcaggc	gacg	gaggct	ttgc
50	101	gccgcg	tgga	cggcAG	TACG	GAggA	aaaaa	tcttcg	tcg	GGCGG	AGAtg
	151	atcgAC	AGGg	accgat	gtt	gcgggA	CaCg	Ttgga	acGTG	TGCGT	GCggg
	201	gtcgtT	ctgG	TTATGG	GTGG	Tggtgg	CA	tC	gATG	ATG	Ttt
	251	TTTCAG	gcac	ttatC	ttCTG	ATGGAC	aatC	AGGGG	CtGAA	TtTCT	TTTTTA
	301	GTTTTg	gcgG	GAGTGT	tggG	CATGa	atacG	ctgAT	GCTGG	CAGTAT	GGtt
55	351	gGCAAC	GTTG	TTCCTG	CGCG	TGAAAG	TGGG	ACGGT	TTTTTC	AGCAGT	CCGG
	401	CGACGT	GTTT	TCGGGG	CAAA	GGCCCT	GTAA	ATCAGG	CGGT	GTTGCG	GCTG
	451	TATGCG	GACC	AGTGGC	GGCA	ACCTTC	GGTA	CGATGG	AAAA	TAGGCG	CAAC
	501	GGCGCA	CAGC	TTGTGG	CTCT	GCACGT	GCT	CGGAAT	GCTG	GTGT	CGGTAT
	551	TGCTG	CTGCT	TTTGGT	GCGG	CAATAT	ACGT	TCAACT	TGGG	AAGCAC	GCTG
	601	TTGAG	CAATG	CCGCTT	CGGT	ACGCGC	GGTG	GAAATG	TTGG	CATGGC	TGCC
60	651	GTCGAA	ACTC	GGTTT	CCCTG	TCCCGA	TGC	GCGGGC	GGTC	ATCGAA	GGTC
	701	GTCTGA	ACGG	CAATAT	TGCC	GATGCG	CGGG	CTTGGT	CGGG	GCTGCT	GGTC
	751	GGCAGT	TATCG	TCTGCT	ACGG	CATCCT	GCCG	CGCCTC	TTGG	CTTGGG	TAGT

WO 99/24578

PCT/IB98/01665

-159-

5
10
1001
1051
1101
1151
1201
1251
1301

```

801 GTGTAAAATC CTTTGTAAAA CAAGCGAAAA CGGattgGAT TTGGAAAAAA
851 CCTATTATCA GCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCcgaaa TCGTCTTGAA
951 CGATGCGCCG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC
1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGCGCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG
1151 TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC
1301 CTGACAGGGT GGCGCAGGAA GGCCGTTTGA AAGACCAATA A

```

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

15
20
401

```

1 MLNPSRKLVE LVRILNKGFF IFSGDPVQAT EALRRVDGST EEKIFRRAEM
51 IDRDRMLRDT LERVAGSFW LWVVASMMF TAGFSGTYLL MDNQGLNFFL
101 VLAGVLGMNT LMLAVWLATL FLRVKVGRRF SSPATWFRGK GPVNQAVLRL
151 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLLVR QYTFNWESTL
201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV
251 GSIVCYGILP RLLAWVCKI LLKTSENGLD LEKTYQAVI RRWQNKITDA
301 DTRRETSAV SPKIVLNDAP KWALMLETEW QDGQWFEGRL AQEWLDKGVA
351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRLQIVR LSEAAQGGAV
401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ*

```

ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

25
30
35
40
45
50
55
60
65

```

                10      20      30      40      50      60
orf33-1.pep    MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDRNRMLRET
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf33ng-1      MLNPSRKLVELVRILNKGFFIFSGDPVQATEALRRVDGSTEEKIFRRAEMIDRDRMLRDT
                10      20      30      40      50      60

                70      80      90      100     110     120
orf33-1.pep    LERVAGSFWLWVVAATFAFFTGFSTYLLMDNQGLNFFLVLAVGLGMNTLMLAVWLAML
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf33ng-1      LERVAGSFWLWVVASMMFTAGFSGTYLLMDNQGLNFFLVLAVGLGMNTLMLAVWLATL
                70      80      90      100     110     120

                130     140     150     160     170     180
orf33-1.pep    FLRVKVGRRFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf33ng-1      FLRVKVGRRFSSPATWFRGKGPVNQAVLRLYADQWRQPSVRWKIGATAHSLWLCTLLGML
                130     140     150     160     170     180

                190     200     210     220     230     240
orf33-1.pep    VSVLLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf33ng-1      VSVLLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
                190     200     210     220     230     240

                250     260     270     280     290     300
orf33-1.pep    DARAWSGLLVGSIACYGILPRLAWVCKILLKTSENGLDLEKPYQAVIRRWQNKITDA
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf33ng-1      DARAWSGLLVGSIIVCYGILPRLAWVCKILLKTSENGLDLEKTYQAVIRRWQNKITDA
                250     260     270     280     290     300

                310     320     330     340     350     360
orf33-1.pep    DTRRETSAVSPKIIILNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVATNREQVALE
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf33ng-1      DTRRETSAVSPKIVLNDAPKWALMLETEWQDGQWFEGRLAQEWLDKGVAAANREQVALE
                310     320     330     340     350     360

                370     380     390     400     410     420
orf33-1.pep    TELKQKPAQLLIGVRAQTVPDRGVLRLQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf33ng-1      TELKQKPAQLLIGVRAQTVPDRGVLRLQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW
                370     380     390     400     410     420

                430     440
orf33-1.pep    RNALAECAAWLEPDRAAQEGRLKDQX
                ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

WO 99/24578

PCT/IB98/01665

-160-

orf33ng-1 RNALTECGAAWLEPDRVAQEGRLKDQX
 430 440

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is
 5 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```

10      1  ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTCGGGCG GCGTGTTCCTT
      51  CGGGGTGTCC GGTCTGGTAT GGTTCCTTTT GGGCGTTTCT TT.GAGTGCG
      101  CCTGTTTTTC GGGTGTTCCT TTTCGGGGTT CGGGACGGGG GACGTTTGTG
      151  GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTG TTCC.GGCGT
      201  CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTTG CG..GTTTGA
      251  CCCGGTTTTT CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
15      301  TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
      351  GGGTTGGGCG GCATCTTGT CCGACTACGC CGTTTGGCAG CCAGAATTCG
      401  GTTTCGCGGG GGCTGTGGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
      451  GTCC..
  
```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```

20      1  ..QKSLSRISLW GLGGVFFGVG GLVWFSLGVS XECACFSGVS FRGSGRGTFV
      51  GSTGVSLSVF SACVXGVVRL PVGLSCVGR LXXLTRFFLGA AGDVILLPLS
      101  SVPSGCAGSD EAAWWCSGWA ASCPTTFFGS QNSVSRGLSV CCGSA*RVLS
      151  S..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

```

25      1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
      51  GCCGGGTCAG AATAGGTTGT CCAGAATTTT TTTATGGGGT TTGGGCGGCG
      101  TGTTCCTTCG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
      151  GGCTGCGCCT GTTTTCGGG TGTTTCTTTT CGGGGTTCGG GACGGGGGAC
      201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTCA GCTTGTGTTC
30      251  CGGCGTCGTC CGGCTGCCTG TCGGTTTGAG CTGTGTCGGC AGGTTGCGGT
      301  TTGACCCGGT TTTTCTTGGG TCGGCGAGG GACGGCAGTC CGCTGCCGCT
      351  TTCGTCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT
      401  GTTCGGGTTG GCGGCGCATCT TGTCCGACTA CGCCGTTTGG CAGCCAGAAT
      451  TCGGTTTCGC GGGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT
35      501  GTCGCCGTTT GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
      551  TGGCGGCGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
      601  AGCCTGAAGG GTTTGTTCGG TTTTGTGACC ATTTTGATTG TGCTTTTGGG
      651  GTGTCGGGCA ATGCCGTCTG AAGGCGGTTT AGACGGCATT GCCGAGTCAG
      701  CGTTGGACGT AGTTTGGTGA GAGGGTGATG ACTTTTGTGA CGCCGACGGT
40      751  GGTGCTGACT TTTTGGGTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC
      801  CCATAACGTA GGTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
      851  GTGGCGGGGC TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTTCCA
      901  AGTGTGCGCG GCGATGTCGC CGGCAGTGG CCGCAGGGAG GCGACGGTAA
      951  TATAGTTGTA CACGCCTTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG
45      1001  ACGAACTGTT TTTCGCCTTC GGTGGCGACT TGTCCGAGCA GCAGCAGGTG
      1051  GCGGTTGTAG CCGACGACGG AGATTTGGGG CGTGTAGCCT TTGGTTTGGT
      1101  TGTTCCTGGC CAGATAGGAA CGGGCGGTGG TTTCGATACG CAACGCCATA
      1151  ACGTTGTCGT CGGTTTGGCG GCCGGTGGTT CGGCGGTCGA CGGCGGATTT
      1201  CGCGCCGACG GCGGCGCTTC CGATTACTGC GCTGACGCAG CCGCTAAGGG
50      1251  CAAGGCTGAA AATGGCGGCA ATCAGGGTGC GGACGGTGTG CGGTTTGGGT
      1301  TTCATCGGGT GCTTCCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC
      1351  CATGCCGTCT GA
  
```

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

```

55      1  MMMPFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLGVS
      51  GCACFSGVSF RSGRGTFVG STGVSLSVFS ACVPASSGCL SV*AVSAGCG
      101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTFFGSQN
      151  SVSRGLSVCC GSA*RVLSPF GLNVLTPMIA NAPMAAIQMS NTARIRSLG
  
```


WO 99/24578

PCT/IB98/01665

-162-

This encodes a protein having amino acid sequence <SEQ ID 212>:

```

1  MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLGVSX
51 SLGV SXGCAC FSGVSFRGSG RGTFVGSTGV SLSVFSACAP ASSGCLSVXA
101 VSAGCGLTRX FXGAAGDGSP LPLSSVPSGC AGADEEAXXC SGWAASCPTT
5  151 PFGSQNSVSR GLSVCCGSVW RVLSPFGXNV LTMPIANAPM AVIQMSNTAR
201 IRSLGVSLKG LFXFFAILIV LLGCRAMPSE GGS DGI AESA LDVVXVEGDD
251 FLYADGGADF LGNLR LFFGG EDAHNVGYVA VGND F DARLC GGADAQQRGA
301 DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL
351 SEQQQVAVVA DNGDLGRVXF GLVVLAQIGA GGGFDTQRHY VVVGXRAGGS
10 401 AVDGGFRADR RAADD CADA AEGKAEDGGS QGADGVRFGE HRVLPFLGVS
451 DGIALRHAV*

```

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

```

15 orf34a.pep 10 20 30 40 50 60
MMXPXIMLPWIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWFSLGVSXSLGV SXGCAC
|| | ||||| ||||| : ||| ||||| ||||| ||||| |||||
orf34-1 M M P F I M L P W I A G V P A V P G Q N R L S R I S L W G L G G V F F G V S G L V W F S L G V S L -----GCAC
10 20 30 40 50

20 orf34a.pep 70 80 90 100 110 120
FSGVSFRGSGRGTFVGSTGVSLSVFSACAPASSGCLSVXAVSAGCGLTRXFXGAAGDGSP
||||| ||||| ||||| : ||||| ||||| ||||| ||||| |||||
orf34-1 FSGVSFRGSGRGTFVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
60 70 80 90 100 110

25 orf34a.pep 130 140 150 160 170 180
LPLSSVPSGCAGADEEAXXC SGWAASCPTT PFGSQNSVSRGLSVCCGSVWRVLS PFGXNV
||||| ||||| : ||| ||||| ||||| ||||| ||||| : ||||| ||
orf34-1 LPLSSVPSGCAGSDEAAW W C SGWAASCPTT PFGSQNSVSRGLSVCCGSAXRVLS PFGGLNV
120 130 140 150 160 170

30 orf34a.pep 190 200 210 220 230 240
LTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSDGIAESA
||||| ||||| : ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf34-1 L T M P I A N A P M A I Q M S N T A R I R S L G V S L K G L F G F F A I L I V L L G C R A M P S E G G S D G I A E S A
180 190 200 210 220 230

35 orf34a.pep 250 260 270 280 290 300
LDVVXVEGDDFLYADGGADFLGNLR LFFGG EDAHNVGYVAVGNDFDARLCGGADAQQRGA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf34-1 L D V V L V E G D D F L Y A D G G A D F L G N L R L F F G G E D A H N V G Y V A V G N D F D A R L C G G A D A Q Q R G A
240 250 260 270 280 290

40 orf34a.pep 310 320 330 340 350 360
DFGCVPSVAGDVAGSARQGGDGNVXVHAFGGLFGTCNLTD ELFLAFGGDLSEQQQVAVVA
||||| ||||| ||||| : ||||| ||||| ||||| : ||||| ||||| |||||
orf34-1 D F G C V P S V A G D V A G S A R Q G G D G N I V V H A F G G L F G T C N L T D E L F F A F G G D L S E Q Q Q V A V V A
300 310 320 330 340 350

45 orf34a.pep 370 380 390 400 410 420
DNGDLGRVXFGLVVLAQIGAGGGFDTQRHYVVVGXRAGGS AVDGGFRADRRRAADD CADA A
|: ||||| ||||| ||||| : ||||| ||||| ||||| ||||| ||||| : |||||
orf34-1 D D G D L G R V A F G L V V L A Q I G T G G G F D T Q R H N V V G L R A G G S A V D G G F R A D G G A S D Y C A D A A
360 370 380 390 400 410

50 orf34a.pep 430 440 450 460
AEGKAEDGGSQGADGVRFGEHRVLPFLGVSDGIALRHAVX
|: ||||| : ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf34-1 A K G K A E N G G N Q G A D G V R F G F H R V L P F L G V S D G I A L R H A V X
420 430 440 450

55 orf34a.pep
orf34-1
60

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from *N. gonorrhoeae*:

orf34.pep

QKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE-----CAC 35

WO 99/24578

PCT/IB98/01665

-164-

		70	80	90	100	110	120
		120	130	140	150	160	170
5	orf34-1.pep	LPLSSVPSGCAGSDEAAWCSGWAASCPTTFFGSQNSVSRGLSVCCGSAXRVLSPFGLNV					
	orf34ng	LPLSSVPSGCAGSDEAAWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLSPFGLNV					
		130	140	150	160	170	180
		180	190	200	210	220	230
10	orf34-1.pep	LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
	orf34ng	LTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
		190	200	210	220	230	240
		240	250	260	270	280	290
15	orf34-1.pep	LDVVLVEGDDEFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
	orf34ng	LDVVLVEGNDFLYADGGADFLGNLRLFFGGEDAHNVGYIAVGNDFDARLCGADAQQRGA					
		250	260	270	280	290	300
		300	310	320	330	340	350
20	orf34-1.pep	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGFLFGTCNLTDLFFAFGGDLSEQQQVAVVA					
	orf34ng	DFGRVPSVAGDVARSAQGGDGNVVVYAFGGFLFGTCNLTDLFFAFGGDLSEQQQVAVVA					
		310	320	330	340	350	360
		360	370	380	390	400	410
30	orf34-1.pep	DDGDLGRVAFGLVVLAAQIGTGGGFDTRHNVVGLRAGGSVDGGFRADGGASDYCADAA					
	orf34ng	DDGDLGRVAFGLVVLAAQVGTGGGFDTRHNVVIGLRAGGSVDGFCADGGPADDCAEAA					
		370	380	390	400	410	420
		420	430	440	450		
35	orf34-1.pep	AKGKAENGGNQGADGVRFGRVLPFLGVSDGIALRHAVX					
	orf34ng	AEGKAEDGGNQGADGVWFGFHRGLPFLGVSDGIALRHAVX					
		430	440	450	460		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

45	1	ATGAAAACCT	TCTTCAAAAC	CCTTTCCGCC	GCCGCACTCG	CGCTCATCCT
	51	CGCCGCCTGC	GGATT.CAAA	AAGACAGCGC	GCCCGCCGCA	TCCGCTTCTG
	101	CCGCCGCCGA	CAACGGCGCG	GCGTAAAAAA	GAAATCGTCT	TCGGCACGAC
	151	CGTCGGCGAC	TTCGGCGATA	TGGTCAAAGA	ACAAATCCAA	GCCGAGCTGG
	201	AGAAAAAAGG	CTACACCGTC	AACTGGTTCG	AGTTTACCGA	CTATGTACGC
	251	CCGAATCTGG	CATTGGCTGA	GGGCGAGTTG		

50 This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

	1	MKTFFKTL	SA AALALILAAC	G.QKDSAPAA	SASAAADNGA	AKKEIVFGTT
	51	VGDFGDMVKE	QIQAELEKKG	YTVKLVEFTD	YVRPNLALAE	GEL

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

55	1	ATGAAAACCT	TCTTCAAAAC	CCTTTCCGCC	GCCGCACTCG	CGCTCATCCT
	51	CGCCGCCTGC	GGCGGTCAAA	AAGACAGCGC	GCCCGCCGCA	TCCGCTTCTG
	101	CCGCCGCCGA	CAACGGCGCG	GCGAAAAAAG	AAATCGTCTT	CGGCACGACC
	151	GTCGGCGACT	TCGGCGATAT	GGTCAAAGAA	CAAATCCAAG	CCGAGCTGGA
	201	GAAAAAAGGC	TACACCGTCA	AACTGGTTCG	GTTTACCGAC	TATGTACGCC

WO 99/24578

PCT/IB98/01665

-165-

5
10
251 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTCCAACAC
301 AAACCCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACC GA
351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
451 CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
501 CAAACTCAAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG
551 CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
851 GCGCAGCCAA ATAA

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

15
20
1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFOH
101 KPYLDDFKKE HNLDTVEFQ VPTAPLGLYP GKLKSLEEVK DGSTVSAPND
151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
201 PRSRADVFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis*:

25
30
35
orf4.pep
orf4a
orf4.pep
orf4a
orf4a

10	20	30	40	50	59
MKTFFKTLA AALALILAACG-QKDSAPAA SASAAADNGA AKKEIVFGTTVGDFGDMVKE					
10	20	30	40	50	60
MKTFFKTLA AALALILAACGGQKDSAPAA SASAAADNGA AXKEIVFGTTVGDFGDMVKE					
60	70	80	90		
QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL					
70	80	90	100	110	120
XIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVXQHXXYLDDXKKXHNLDITXVXQ					
130	140	150	160	170	180
VPTAPLGLYPGKLKSLXXVKXGSTVSAPNDPXXFXRVLVMLDELGXIKLKDXIXXXXXXX					

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

40
45
50
55
1 ATGAAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
51 CGCCGCCTGC GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGAANAAAG AAATCGTCTT CGGCACGACC
151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA
201 GAAAAAAGGC TACACCGTCA AACTGGTCTG GTNTACCGAC TATGTGCGCN
251 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTNCAACAC
301 ANACNCTATC TTGACGACTN CAAAAANAA CACAATCTGG ACATCACCNN
351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC
451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT
501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN
551 NNNNANNNNT NNNNNNNNN NNNNNCNCG NNNNNNNANN NNNNNNNNN
601 NCGNNTNNNN NNGCNNNNNT NNANNTNNN NNCNNCNNNN NNNNTNNNN
651 NANNANNAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
851 GCGCAGCCAA ATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AXKEIVFGTT

WO 99/24578

PCT/IB98/01665

-166-

51 VGDFGDMVKE XIQPELEKKG YTVKLVEXTD YVRXNLALAE GELDINVXQH
 101 XXYLDDXKKX HNLDITXVXQ VPTAPLGLYP GKLKSLXXVK XGSTVSAPND
 151 PXXFXRVLM LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX
 201 XXXXAXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

1 ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
 51 CGCCGCCTGC GGCGGTCAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
 101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
 151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
 201 GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTGCGCC
 251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC
 301 AAACCCTATC TTGACGACTT CAAAAAAGAA CACAATCTGG ACATCACCAG
 351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
 401 AATCGCTGGA AGAAGTCAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
 451 CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
 501 CAAACTCAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
 551 CCGAAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
 601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
 851 GCGCAGCCAA ATAA

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
 51 VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFOH
 101 KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLKSLEEVK DGSTVSAPND
 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLNKI KIVELEAAQL
 201 PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

35	orf4a-1	10 20 30 40 50 60	MKTFFKTLA AALALILAACGGQKDSAPAA SASAAADNGA AKKEIVFGTTVGDFGDMVKE
	orf4-1	10 20 30 40 50 60	MKTFFKTLA AALALILAACGGQKDSAPAA SASAAADNGA AKKEIVFGTTVGDFGDMVKE
40	orf4a-1	70 80 90 100 110 120	QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFOH KPYLDDFKKE HNLDITEVFQ
	orf4-1	70 80 90 100 110 120	QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFOH KPYLDDFKKE HNLDITEVFQ
45	orf4a-1	130 140 150 160 170 180	VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
	orf4-1	130 140 150 160 170 180	VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
50	orf4a-1	190 200 210 220 230 240	ADIAENLNKI KIVELEAAQL PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS
	orf4-1	190 200 210 220 230 240	ADIAENLNKI KIVELEAAQL PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS
55	orf4a-1	250 260 270 280	AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX
	orf4-1	250 260 270 280	AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX

WO 99/24578

PCT/IB98/01665

-167-

Homology with an outer membrane protein of *Pasteurella haemolitica* (accession q08869).

ORF4 and this outer membrane protein show 33% aa identity in 91aa overlap:

```

5      lip2.pasha      MNFKLLGVALVSALALTACKDEKAQAP----
      ORF4      VXTPNPDGRTPCPSFLFETATTSGENMKTFEKTLSAAAL--ALILAACGFKKTARPPHPL
              110      120      130      140      150

10     lip2.pasha      30      40      50      60      70      80
      lip2.pasha      -ATTAKTENKAPLKVGVMTEGPEAQMTEVAVKIAKEKYGLDVELVQFTEYTPNAALHSKD
      ORF4      LPPPTTARRKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGE
              160      170      180      190      200      210

15     lip2.pasha      90      100      110      120      130      140
      lip2.pasha      LDANAFQTVPYLEQEVKDRGYKLAIIGNTLVWPPIAAYSKKIKNISELKDGATVAIPNNAS
      ORF4      L.....

```

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae*:

```

25     orf4nm.pep      10      20      30
      orf4ng      RANAVXTPNPDGRTPCLSFLFETATTSGENMKTFEKTLSAAALALILAACGXQKDSAPAA
              200      210      220      230      240      250

30     orf4nm.pep      40      50      60      70      80      89
      orf4ng      SASA-AADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA
              260      270      280      290      300      310

35     orf4nm.pep      90
      orf4ng      EGEL
              320      330      340      350      360      370

```

40 The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

```

45      1  MKTFFKTLST  ASLALILAAC  GGQKDSAPAA  SAAAPSADNG  AAKKEIVFGT
      51  TVGDFGDMVK  EQIQAELEKK  GYTVKLVEFT  DYVRPNLALA  EGELDINVFO
      101 HKPYLDDFKK  EHNLDITEAF  QVPTAPLGly  PGKLKSLEEV  KDGSTVSAPN
      151 DPSNFARALV  MLNELGWIKL  KDGINPLTAS  KADIAENLKN  IKIVELEAAQ
      201 LPRSRADVDF  AVVNGNYAIS  SGMKLTEALF  QEPSFAYVNW  SAVKTADKDS
      251 QWLKDVTEAY  NSDAFKAYAH  KRFEGYKYP  A  AWNEGAAK*

```

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

```

50      1  atgAAAACCT  TCTTCAAAAC  cctttccgcc  gccgcaCTCG  CGCTCATCCT
      51  CGCAGCCTGc  ggCggtcaAA  AAGACAGCGC  GCCCgcagcc  tctgcCGCCG
      101 CCCCTTCTGC  CGATAACGgc  gCgGCGAAAA  AAGAAAtcgt  ctTCGGCACG
      151 Accgtgggcg  acttcggcgA  TAtggTCAAA  GAACAAATCC  AagcCGAgct
      201 gGAGAAAAAA  GgctACACcg  tcAAattggt  cgaatttacc  gactatgtGC
      251 gCCCGAATCT  GGCATTGGCG  GAGGGCGAGT  TGGACATCAA  CGTCTTCCAA
      301 CACAAACCTT  ATCTTGACGA  TTTCAAAAAA  GAACACAACC  TGGACATCAC
      351 CGAAGCCTTC  CAAGTGCCGA  CCGCGCCTTT  GGGACTGTAT  CCGGGCAAAC
      401 TGAAATCGCT  GGAAGAAGTC  AAAGACGGCA  GCACCGTATC  CGCGCCCAac
      451 gACccgTCCA  ACTTCGCACG  CGCCTTGGTG  ATGCTGAACG  AACTGGGTTG
      501 GATCAAACCT  AAAGACGGCA  TCAATCCGCT  GACCGCATCC  AAAGCCGACA
      551 TCGCGGAAAA  CCTGAAAAAC  ATCAAAATCG  TCGAGCTTGA  AGCCGCACAA

```

WO 99/24578

PCT/IB98/01665

-168-

5
 601 CTGCCGCGCA GCCGCGCCGA CGTGGATTTT GCCGTCGTCA ACGGCAACTA
 651 CGCCATAAGC AGCGGCATGA AGCTGACCGA AGCCCTGTTC CAAGAGCCGA
 701 GCTTTGCCTA TGTCAACTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC
 751 CAATGGCTTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC
 801 CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG
 851 AAGGCGCAGC CAAATAA

This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:

10
 1 MKTFFKTLA AALALILAAC GGQKDSAPAA SAAAPSADNG AAKKEIVFGT
 51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ
 101 HKPYLDDEFK EHNLDITEAF QVPTAPLGLY PGKLKSLLEV KDGSTVSAPN
 151 DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ
 201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
 251 QWLKDVTEAY NSDAFKAYAH KRFEGYKYP AAWNEGAAK*

This shows 97.6% identity in 288 aa overlap with ORF4-1:

15
 orf4-1.pep 10 20 30 40 50 59
 MKTFFKTLA AALALILAACGGQKDSAPAA SAAAPSADNG AAKKEIVFGT TVGDFGDMVK
 orf4ng-1 MKTFFKTLA AALALILAACGGQKDSAPAA SAAAPSADNG AAKKEIVFGT TVGDFGDMVK
 20
 orf4-1.pep 60 70 80 90 100 110 119
 EQIQAELEKKGYTVKLVEFTDYVRPNLALA EGELDINVFQ HKPYLDDEFK EHNLDITEVF
 orf4ng-1 EQIQAELEKKGYTVKLVEFTDYVRPNLALA EGELDINVFQ HKPYLDDEFK EHNLDITEAF
 25
 orf4-1.pep 120 130 140 150 160 170 179
 QVPTAPLGLYPGKLKSLLEV KDGSTVSAPNDPSNFARVLV MLDELGWIKL KDGINPLTAS
 orf4ng-1 QVPTAPLGLYPGKLKSLLEV KDGSTVSAPNDPSNFARVLV MLDELGWIKL KDGINPLTAS
 30
 orf4-1.pep 180 190 200 210 220 230 239
 KADIAENLKN IKIVELEAAQLPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW
 orf4ng-1 KADIAENLKN IKIVELEAAQLPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW
 35
 orf4-1.pep 240 250 260 270 280
 SAVKTADKDS QWLKDVTEAY NSDAFKAYAH KRFEGYKSP AAWNEGAAKX
 orf4ng-1 SAVKTADKDS QWLKDVTEAY NSDAFKAYAH KRFEGYKSP AAWNEGAAKX
 40
 orf4-1.pep 240 250 260 270 280
 SAVKTADKDS QWLKDVTEAY NSDAFKAYAH KRFEGYKSP AAWNEGAAKX
 orf4ng-1 SAVKTADKDS QWLKDVTEAY NSDAFKAYAH KRFEGYKSP AAWNEGAAKX

45 In addition, ORF4ng-1 shows significant homology with an outer membrane protein from the database:

50
 ID LIP2_PASHA STANDARD; PRT; 276 AA.
 AC Q08869;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE 28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR. . . .
 SCORES Init1: 279 Initn: 416 Opt: 494
 Smith-Waterman score: 494; 36.0% identity in 275 aa overlap
 55
 orf4ng-1.pep 10 20 30 40 50
 MKTFFKTLA AAL--ALILAACGGQKDSAPAA SAAAPSADNG AAKKEIVFGT TVGDFGDM
 lip2_pasha MNFKLLGVALVSALALTACKDEKAQAPATTA---KTENKAPLK---VGVMTGPEAQM
 60
 orf4ng-1.pep 60 70 80 90 100 110
 VKEIQAELEKKGYTVKLVEFTDYVRPNLALA EGELDINVFQ HKPYLDDEFK EHNLDITE
 :: :: || | :||:|:|:|:| || :|| |:| |||:| |::: ::

WO 99/24578

PCT/IB98/01665

-169-

	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTPQNAALHSDKLDANAFQTVPYLEQEVKDRGYKLAI	60	70	80	90	100	110
5	orf4ng-1.pep	AFQVPTAPLGLYPGKLKSLKEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLT	120	130	140	150	160	170
	lip2_pasha	IGNTLVWPIAAYSKKIKNISELKDGATVAIPNNASNTARALLLQAHGLLKLKDKPN-VF	120	130	140	150	160	170
10	orf4ng-1.pep	ASKADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA	180	190	200	210	220	230
	lip2_pasha	ATENDIIEENPKNIKIVQADTSLLTRMLDDVELAVINNTYAGQAGLSPDKDGIIVESKDSP	180	190	200	210	220	230
15	orf4ng-1.pep	YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAX	240	250	260	270	280	289
20	lip2_pasha	YVNLVVSREDNKDDPRLQTFVKSQTEEVFQELKLFNGGVVKGW	240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolitica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
101	GTTTTATGGC	TGAACCAGCA	TTATTTCCAC	GGCAACCTCC	TCTTCTACCT
40	151	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG	CTGGGCGGCG
	201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT
	251	GACAACGGCA	GCGAATGGCT	CGACAGCGGA	CTCATGCGCG
	301	CCTCATCGGC	GyGGCCATCG	CCATCGCCGC	CGCCAAACTG
	351	AATCCACACT	GATGTGGCGT	TTCATGCTTG	CCGACAACCT
45	401	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC	AGGCGCATGA
	451	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA	AATCAACGCA
	501	AAAGCCGCAG	CCATCTCGCC	GCCACATCGG	GCGAAAGCTG
	551	GCCATGATGG	AAGCCATGCA	GCACGCCAC	CGTAAATCG
	601	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT	GCAATCTCCC

55

1	MDRDDRLLRP	RHAPVPRDL	LQGGTYARY	GHRAGRGFR	FMAEPALFPR
51	QPPLLPDHRH	GKRTGRLGGG	RQKRLRPYVG	GADDVHAHRR	QRQRMARQRP
101	DARDERPHER	RHRHCRRTA	AAEIHTDVAF	HACRQPGRLQ	QNDCRNQORQ
151	AYDARTFGAE	YGQAPNORT	HGQKPQPPR	HIGRKPHQPL	HDGSHAARPP

WO 99/24578

PCT/IB98/01665

-171-

201 QNRQHHRAAP DHRROAAISQ TQRQRNPAAR PPLHTAPNRP ATNRRPHQRQ
 251 TRPPHPHRHR HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH
 301 PPQMAGCPRT PTPAPKPA*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

```

10      1  ..GAAATCAGCC TGCGGTCCGA CNACAGGCCG GTTTCCTGN CGAAGCGGCG
      51  GGATTCGGAA CGTTTTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT
      101  GGGCGTGGGT GGAAAACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
      151  CGCGATTTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
      201  TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAA AAGGCACAAG
      251  TGCAGGAACA GTCGCCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
      15  301  GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
      351  CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTAGCCGC AACGCCTGCG
      401  TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
      451  GGACATTATC TCGGAGA.GG AACCATCATG CCCGGTTTCC ACCTGATGAA
      501  AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGGTAAGC
      20  551  GTTATCCTTT CCGACCGG..

```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```

      1  ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKWAWVENG TFATVGSAPY
      51  RDLSPGLAEW AEKADGNVRI VGCAVCGEFK KAQVQEQLAR KIEWLPSSAQ
      101  AXGIRNHYRH PEEHGSDFWF NALGSRFRSR NACVVVSCGT AVTVDALTDD
      25  151  GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..

```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

```

      1  ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
      51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
      101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATA CGCGGG
      30  151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
      201  TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCT GGTTTTCAGA
      251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
      301  GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
      35  351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCT CACCGTTTGG
      401  GCGAGTGTCT GATGTTTCTG TTTGGCTGGG TGTTTGACCG GCCGCGATAT
      451  GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTCTC GGCGCGCCTT
      501  GTCGCGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTGTTGTTG
      551  TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
      601  GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
      40  651  GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCTC ACGGCATCGC
      701  GGCGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAACGCT GTTGGTGGAA
      751  CTGGACGCGG TGTTGTTGCA ATATGCGCGG GACGATTTG CGCCTTTTGT
      801  GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
      851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAG CGTGGACGGA
      45  901  CAAGGCGTTT TGCATTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
      951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
      1001  GGCGGGATTG GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
      1051  AAGTGGGCGT GGGTGGAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
      1101  GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
      50  1151  GAAATGTCCG CATCGTCGGT TGCGCTGTGT GCGGAGAATT CAAAAAGGCA
      1201  CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
      1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
      1301  CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
      1351  TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCAG
      55  1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
      1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
      1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCA GCGGCATGAT
      1551  GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
      1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG

```

```

40      orf61.pep      EISLRSDXRPVSVXKRRDSERFLLLDGGNS
                        ||||||| ||||| ||||||| |||||||
      orf61a      TVFEGTVKGV DGGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNS
      290          300          310          320          330          340

45      orf61.pep      RLKWAWVENGT FATVGSAPYRDL SPLGAEWA EKADGNVRIVGCAVCGEFKKAQVQEQLAR
                        ||||||| ||||||| ||||||| ||||||| : ||||||| ||||||| ||||||| |||||||
      orf61a      RLKWAWVENGT FATVGSAPYRDL SPLGAEWA EKVDGNVRIVGCAVCGEFKKAQVQEQLAR
      350          360          370          380          390          400

50      orf61.pep      KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTTDD
                        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
      orf61a      KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTTDD
      410          420          430          440          450          460

55      orf61.pep      GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT
                        ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
      orf61a      GHYLG-GTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMM
      470          480          490          500          510          520

60      orf61a      HGRLKEKTGAGKPV DVIITGGGA AKVAEALPPAFLAENTVRVADNLVIHGLLNLI A AEGG

```

WO 99/24578

PCT/IB98/01665

-173-

530 540 550 560 570 580

The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```

1  ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
5  101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CACACGCGGG
151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
201  TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301  GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGTG TGACCCACCT
10 351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401  GCGAGTGTCT GATGTTTCACT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
451  GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGCC GGCGCGCCTT
501  GTCGCGTTTG GGTGTTGAAA CGCAAATCAA GTGGCCAAAC GATTGTTGTCG
551  TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
15 601  GGCAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTCG TGCTGCCCAA
651  GGAAGTGAA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
701  GGCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAACGCT GTTGGCGGAA
751  CTTGATGCGG TGTGTTGCA ATATGCGCGG GACGGATTTG CGCCTTTTGT
801  GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
20 851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
901  CAAGGCGTTC TGCACTTGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGTTTCC GTGCCGAAGC
1001 GGCGGGATTG GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGGTGGAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
25 1101 GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGTGGATG
1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATT CAAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGCGAGC GCCGCTTCAG CCGCAACGCC
30 1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGGG GAACATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501 CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTGCGCA GCGGCATGAT
1551 GGATGCGGTT TCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
35 1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCAAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```

40 1  MTVLKPSHWR VLAELADGLP QHVSQARMAD MKPQQLNGF WQQMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLRELTERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDPRQY
151 ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGRDKLG GILLETVRTG
201 GKTVAUVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
45 251 LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVDG
301 QGVLEHLETA GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
351 KWAUVENGTF ATVGSAPYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA
401 QVQEQRLARKI EWLPSAQAAL GIRNHYRHPE EHGSDFWFNA LGSRRFSRNA
451 CVVVSCTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANLNRHAGK
50 501 RYFPFTTTGN AVASGMMDAV CGSVMMMHGR LKEKTGAGKP VDVIIITGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

```

ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```

55 orf61a.pep 10 20 30 40 50 60
MTVLKPSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR
|||||
orf61-1 MTVLKLSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR
10 20 30 40 50 60

60 orf61a.pep 70 80 90 100 110 120
LVRPLAVFDAEGLRELTERS GFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
|||||
orf61-1 LVRPLAVFDAEGLRELTERS GFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
70 80 90 100 110 120

65 130 140 150 160 170 180

```

WO 99/24578

PCT/IB98/01665

-174-

5	orf61a.pep	GRGRQGRKWSHRLGECLMFSFGWVFD RPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVFD RPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN
10	orf61a.pep	DLVVGRDKLGGIL IETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA
	orf61-1	DLVVGRDKLGGIL IETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA
15	orf61a.pep	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG
20	orf61a.pep	QGV LHLETAEGKQTVVS GEISLRSDDRPVSVPKRRDSE RFLLLDGGNSRLKWAWVENGT F
	orf61-1	QGV LHLETAEGKQTVVS GEISLRSDDRPVSVPKRRDSE RFLLLDGGNSRLKWAWVENGT F
25	orf61a.pep	ATVGSAPYRDLSP LGAEWA EKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL
	orf61-1	ATVGSAPYRDLSP LGAEWA EKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL
30	orf61a.pep	GIRNH YRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTV DALTD DGHYLG GTIMP GF
	orf61-1	GIRNH YRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTV DALTD DGHYLG GTIMP GF
35	orf61a.pep	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP
	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP
40	orf61a.pep	VDVIITGGGA AKVAEALPPAFLAENTVRVADNLVIHGLLNLI AEGGESEHTX
	orf61-1	VDVIITGGGA AKVAEALPPAFLAENTVRVADNLVIYGLNMI AEGGREYEHIX

Homology with a predicted ORF from *N.gonorrhoeae*ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*50 *gonorrhoeae*:

55	orf61.pep	EISLRSDXRPVSVXKRRDSE RFLLLDGGNS	30
	orf61ng	TVCEGTVKGV DGRGV LHLETAEGEQTVVS GEISLRPDNRSVSVPKRPDSE RFLLLDGGNS	211
60	orf61.pep	RLKWAWVENGT FATVGSAPYRDLSP LGAEWA EKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKWAWVENGT FATVGSAPYRDLSP LGAEWA EKADGNVRIVGCAVCGESKKAQVKEQLAR	271
65	orf61.pep	KIEWLPSSAQAXGIRNH YRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTV DALTD D	150
	orf61ng	KIEWLPSSAQALGIRNH YRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTV DALTD D	331
	orf61.pep	GHYLGXGTIMP GFHLMKESLAVRTANLNRHAGKRYPFPT	189
	orf61ng	GHYLG-GTIMP GFHLMKESLAVRTANLNRPA GKRYPFPTTTGNAVASGMMDAVCGS IMM	390

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET
COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2

NOTE: Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE
THAN ONE VOLUME

THIS IS VOLUME OF

NOTE: For additional volumes please contact the Canadian Patent Office

PCT/IB98/01665 – AMENDED CLAIMS (Article 19 PCT)

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 458, 460, 462, 464 and 466.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 457, 459, 461, 463 and 465.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 10 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 15 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 20 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 25 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 30 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892.
5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

7. An antibody which binds to a protein according to any one of claims 4 to 6.

8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.

9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
 5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
 10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
 15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
 & 891..

10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the
 group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
 20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
 25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
 30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,
 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,
 35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

WO 99/24578

PCT/IB98/01665

-498-

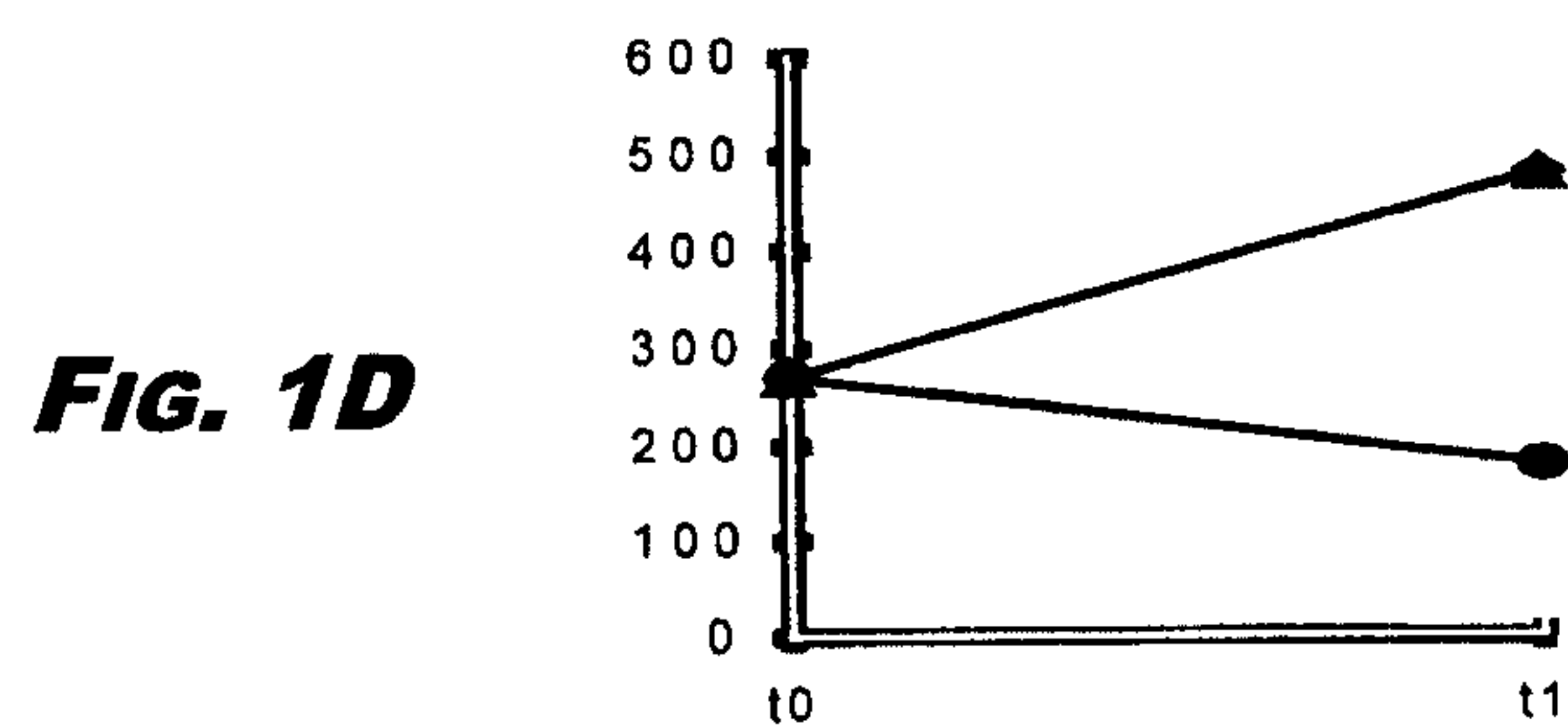
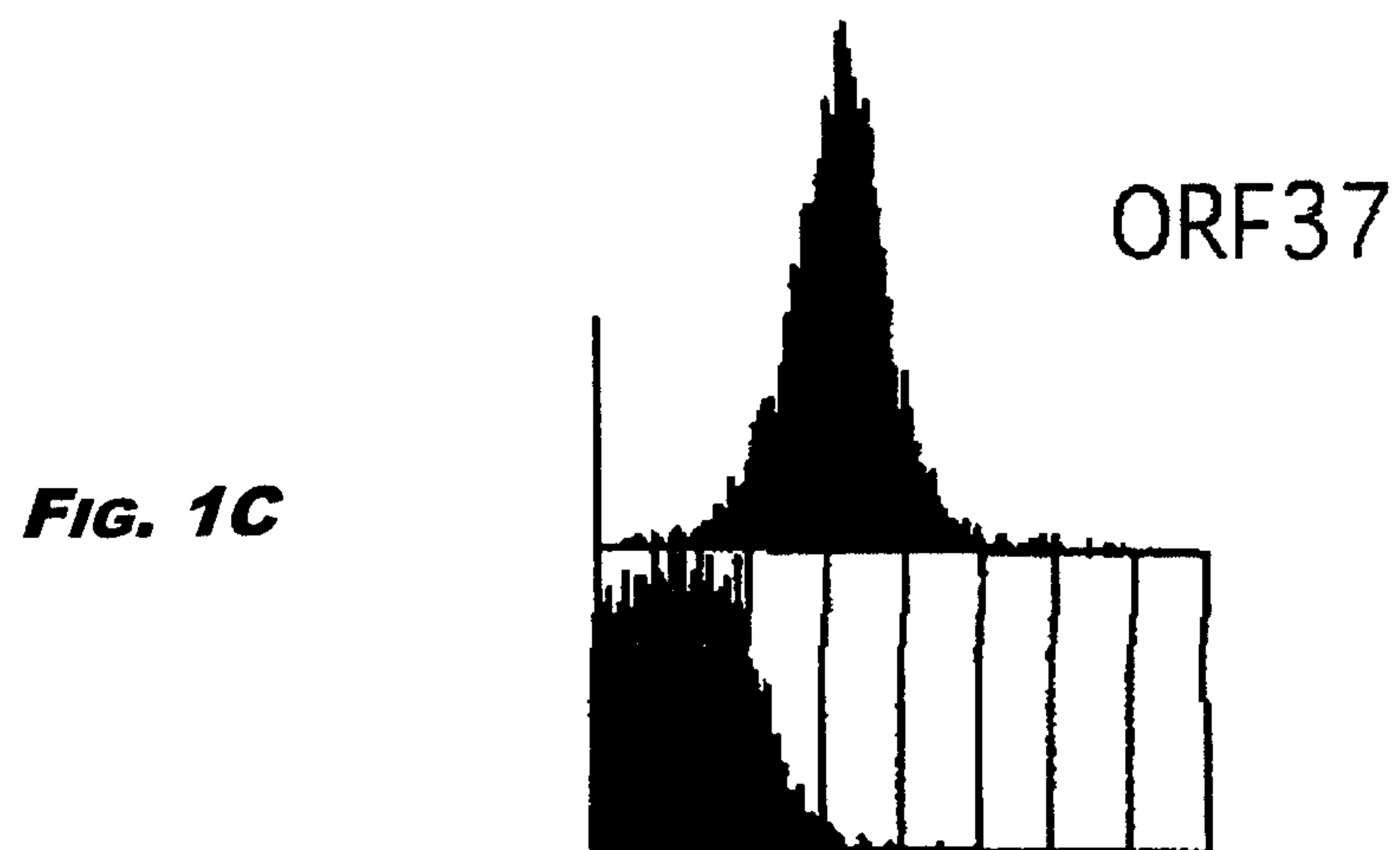
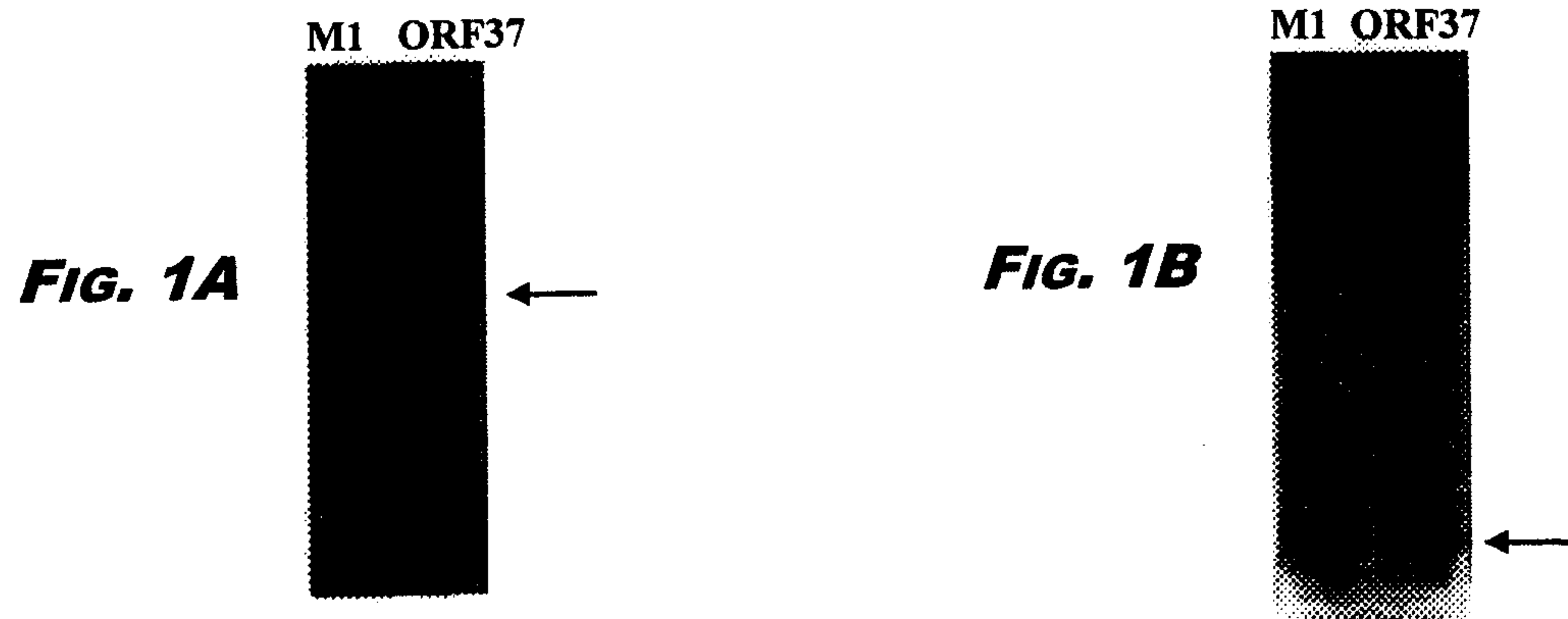
735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
- 10 13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

WO 99/24578

PCT/IB98/01665

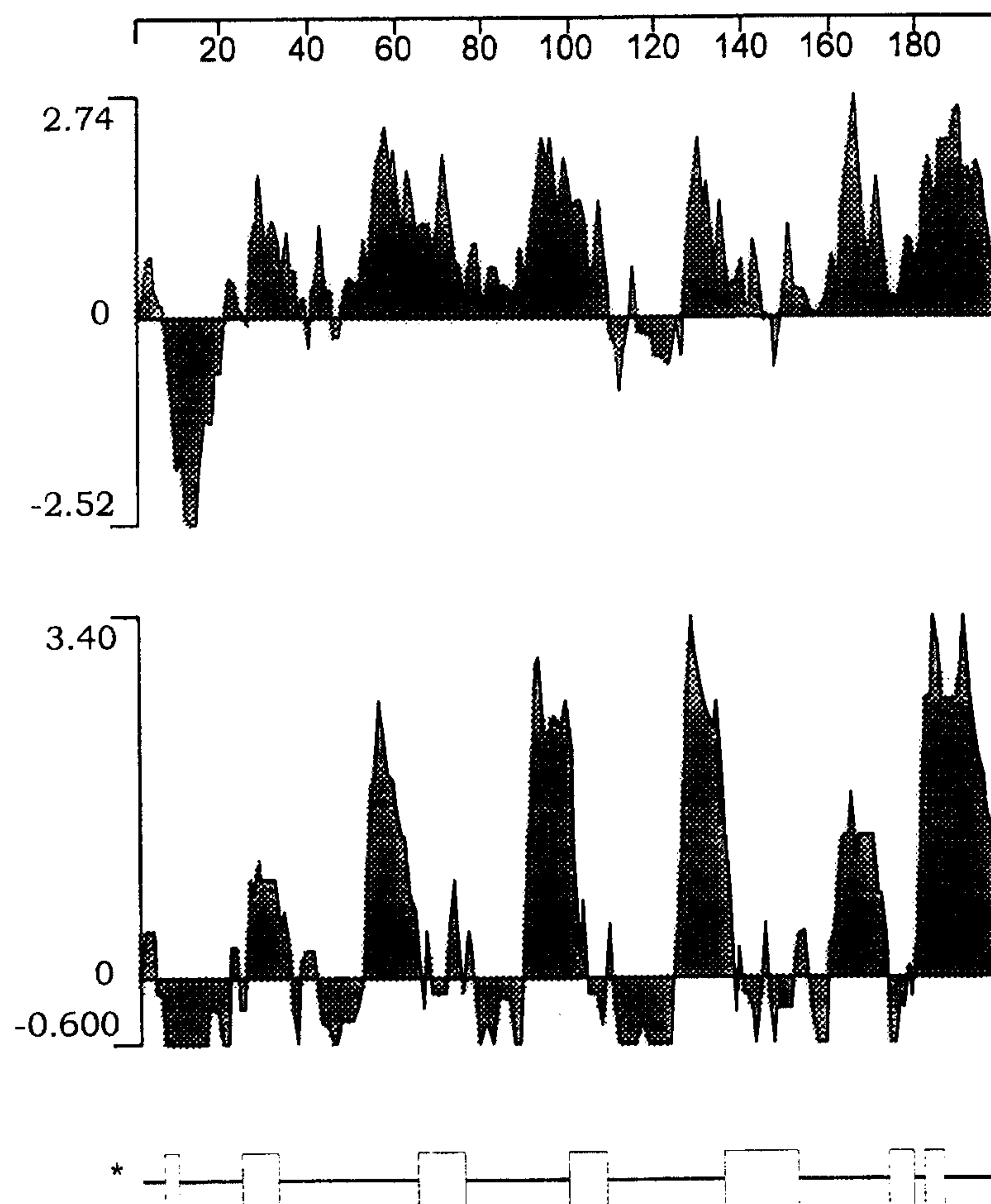
1/24

FIGURE 1

WO 99/24578

PCT/IB98/01665

2/24

Fig 1E

WO 99/24578

PCT/IB98/01665

3/24

FIGURE 2

Fig. 2A

M1 ORF5

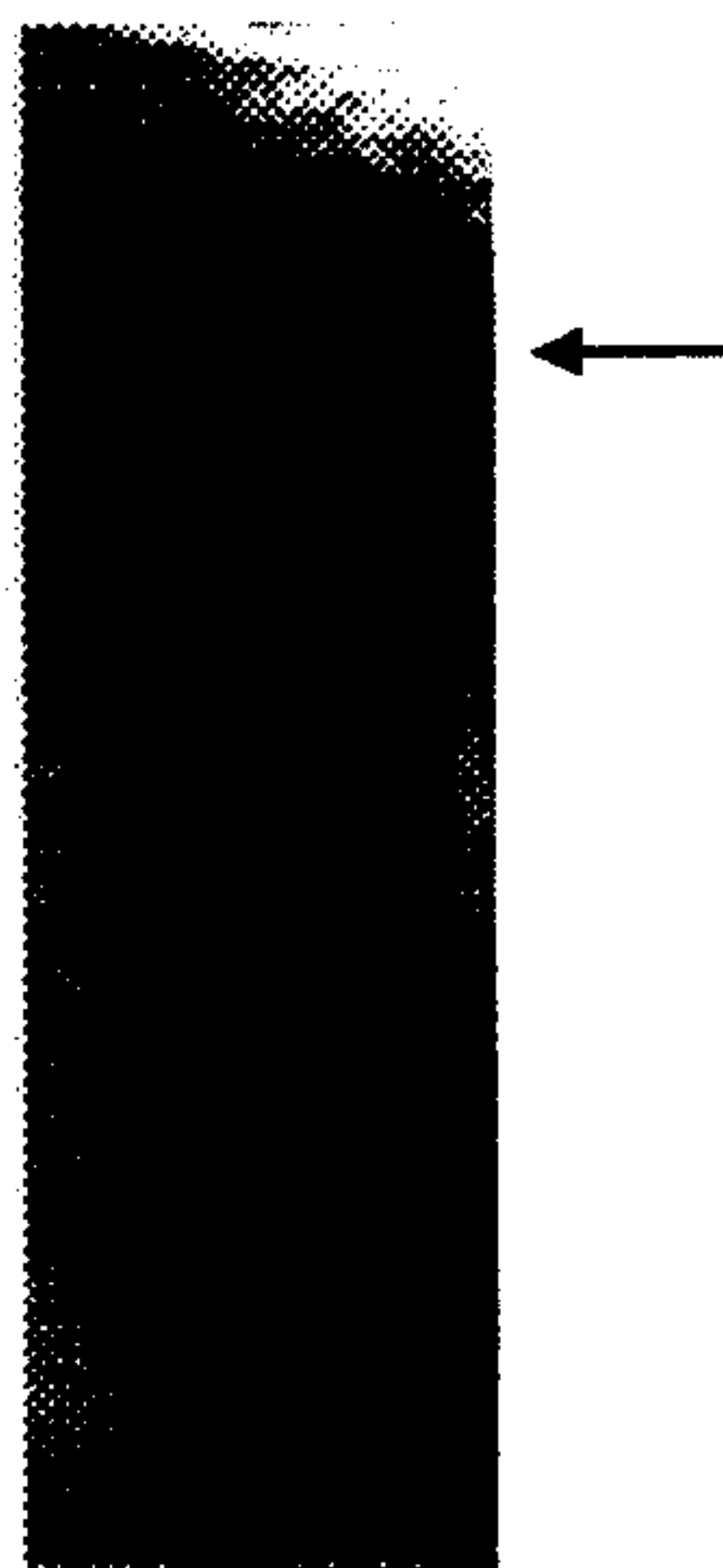


Fig. 2B

TP



FIGURE 3

Fig. 3A

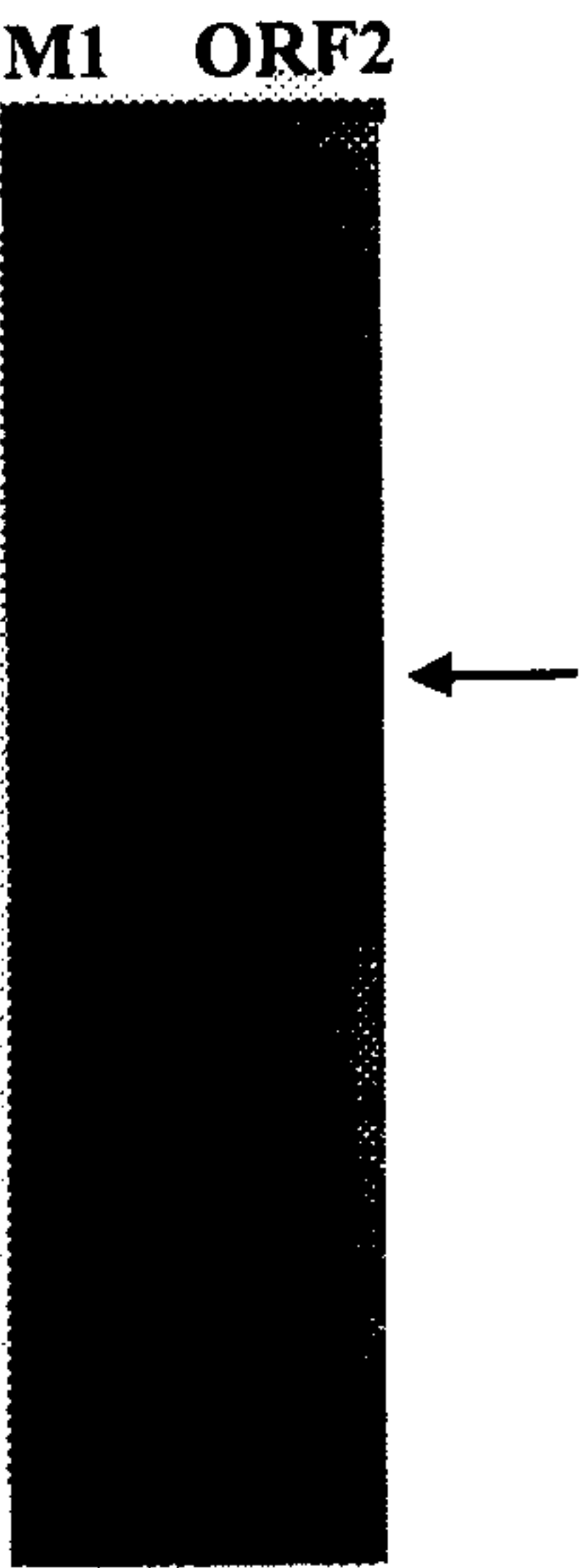


Fig. 3B

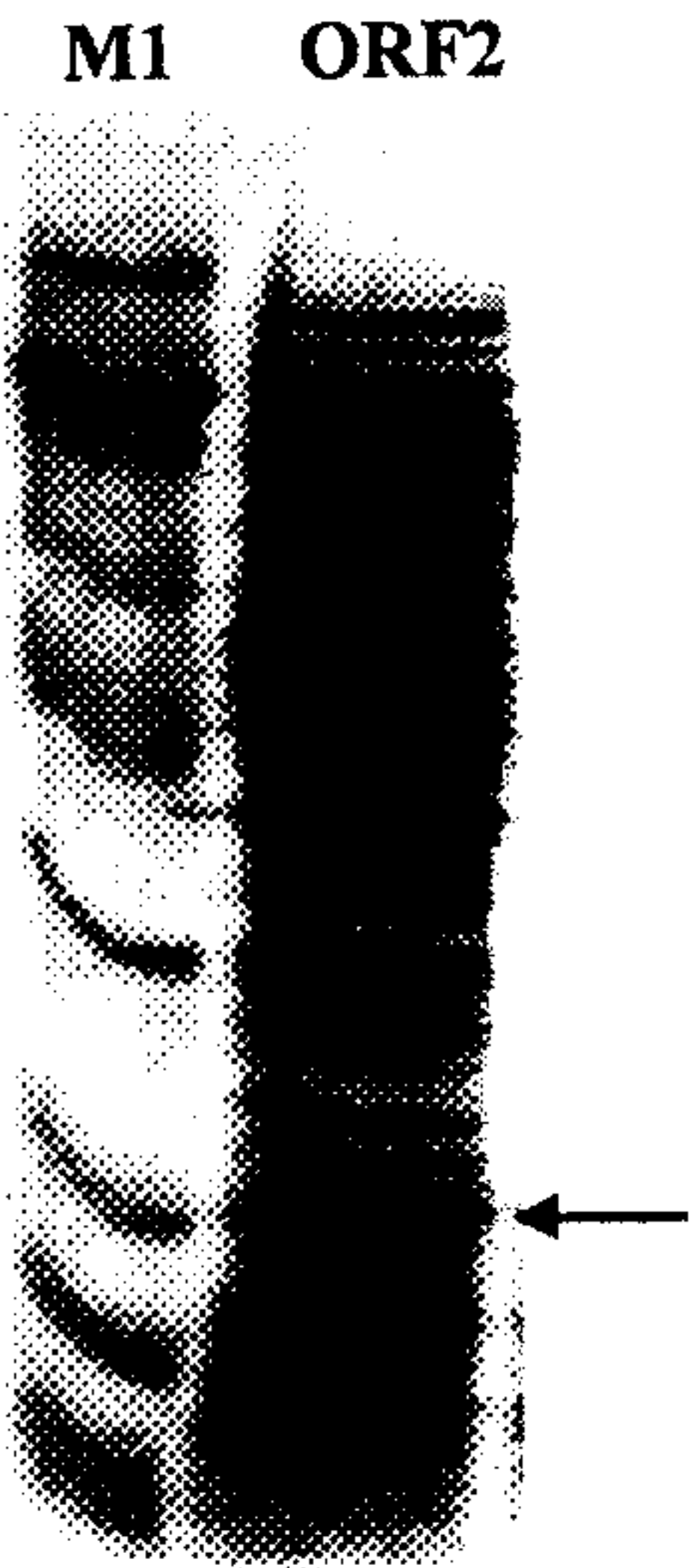


Fig. 3C

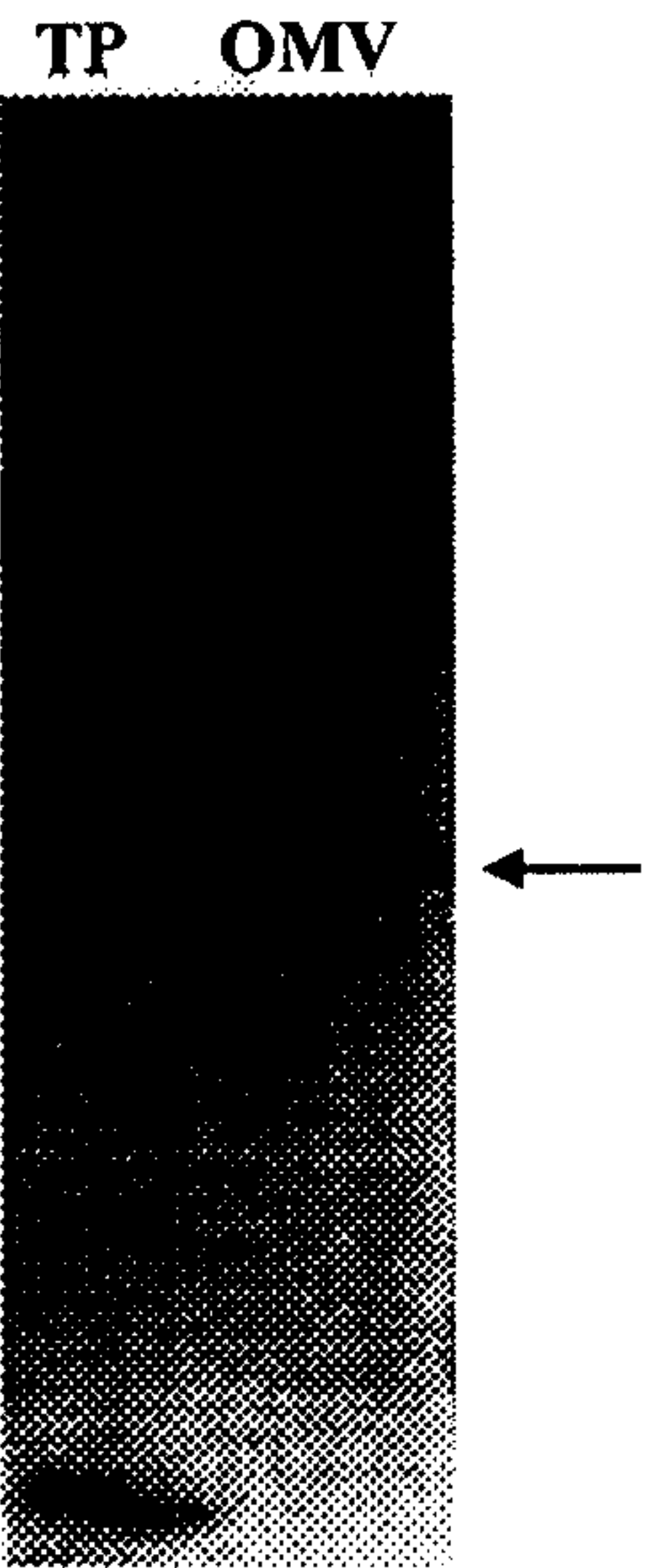
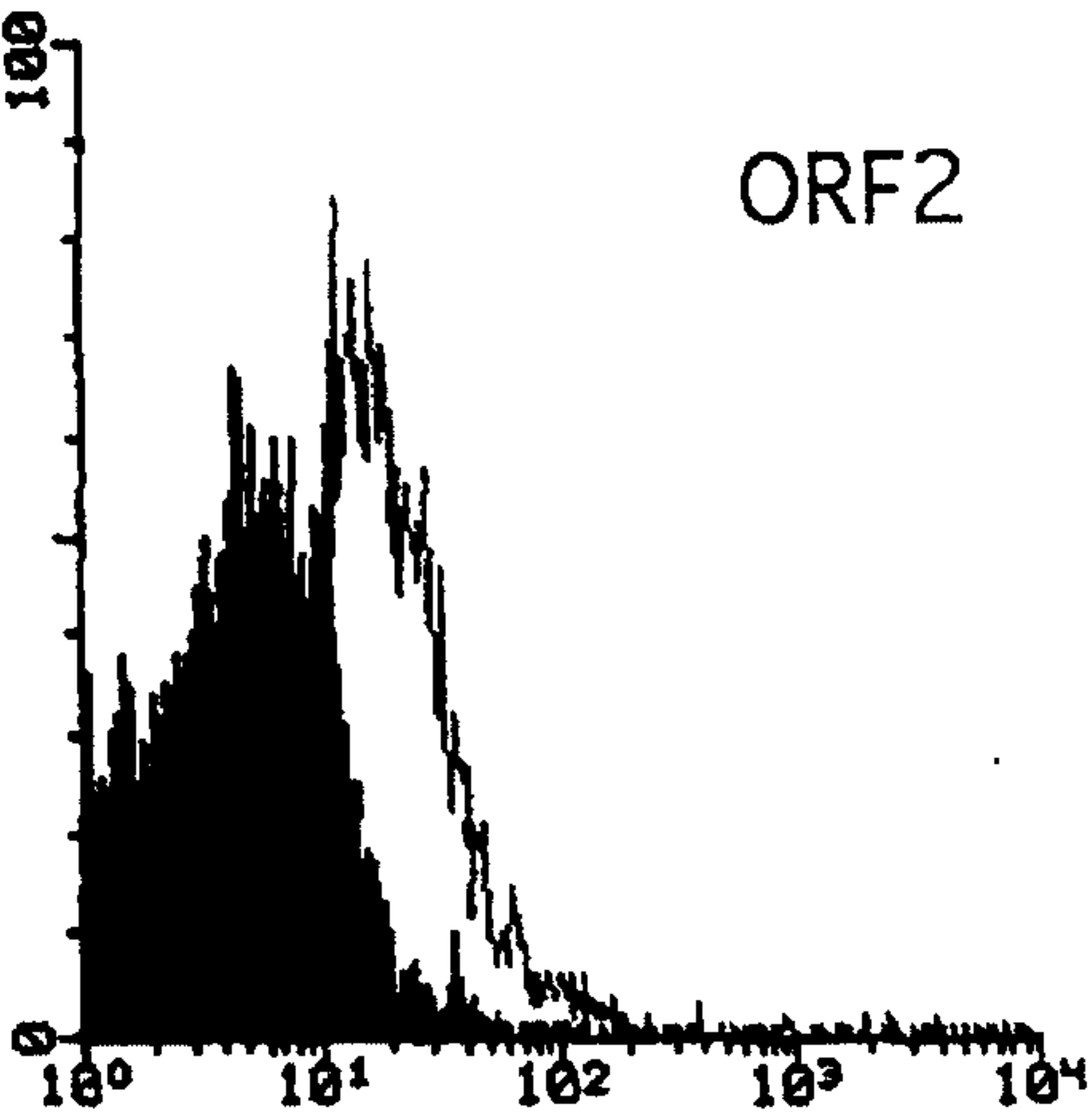


Fig. 3D



5/24

FIGURE 4

Fig. 4A

M1 ORF15



Fig. 4B

M2 ORF15

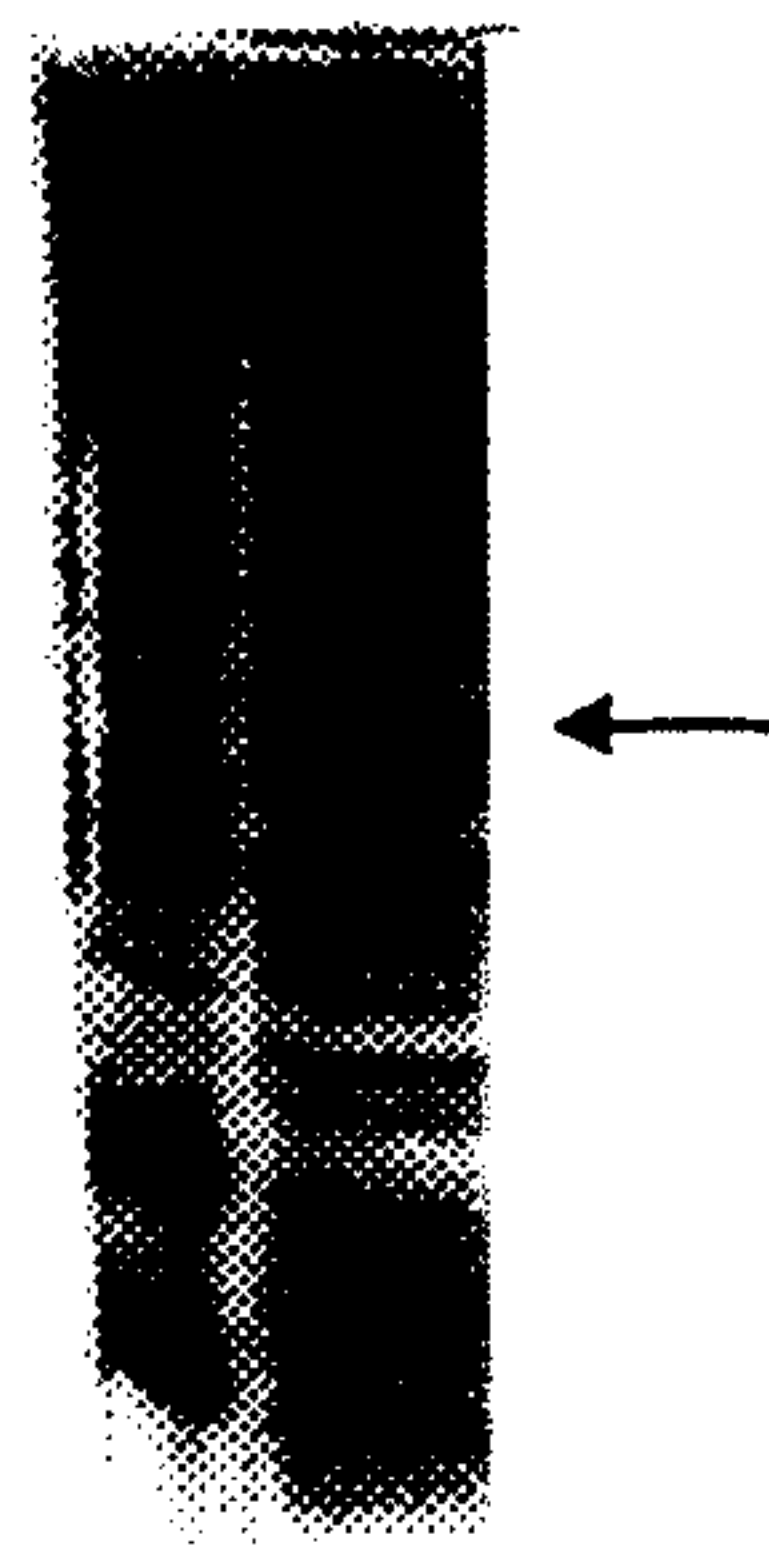
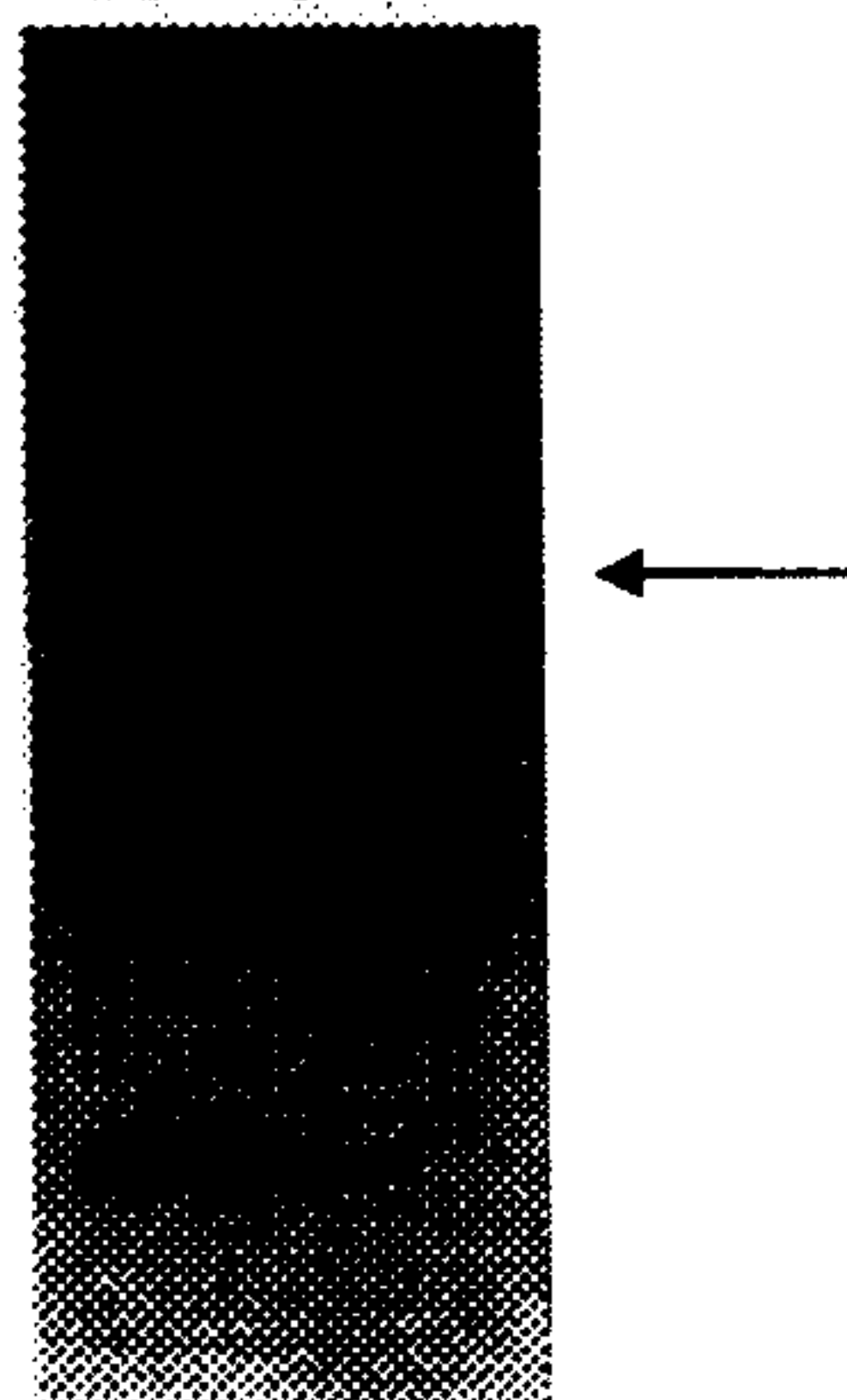


Fig 4C

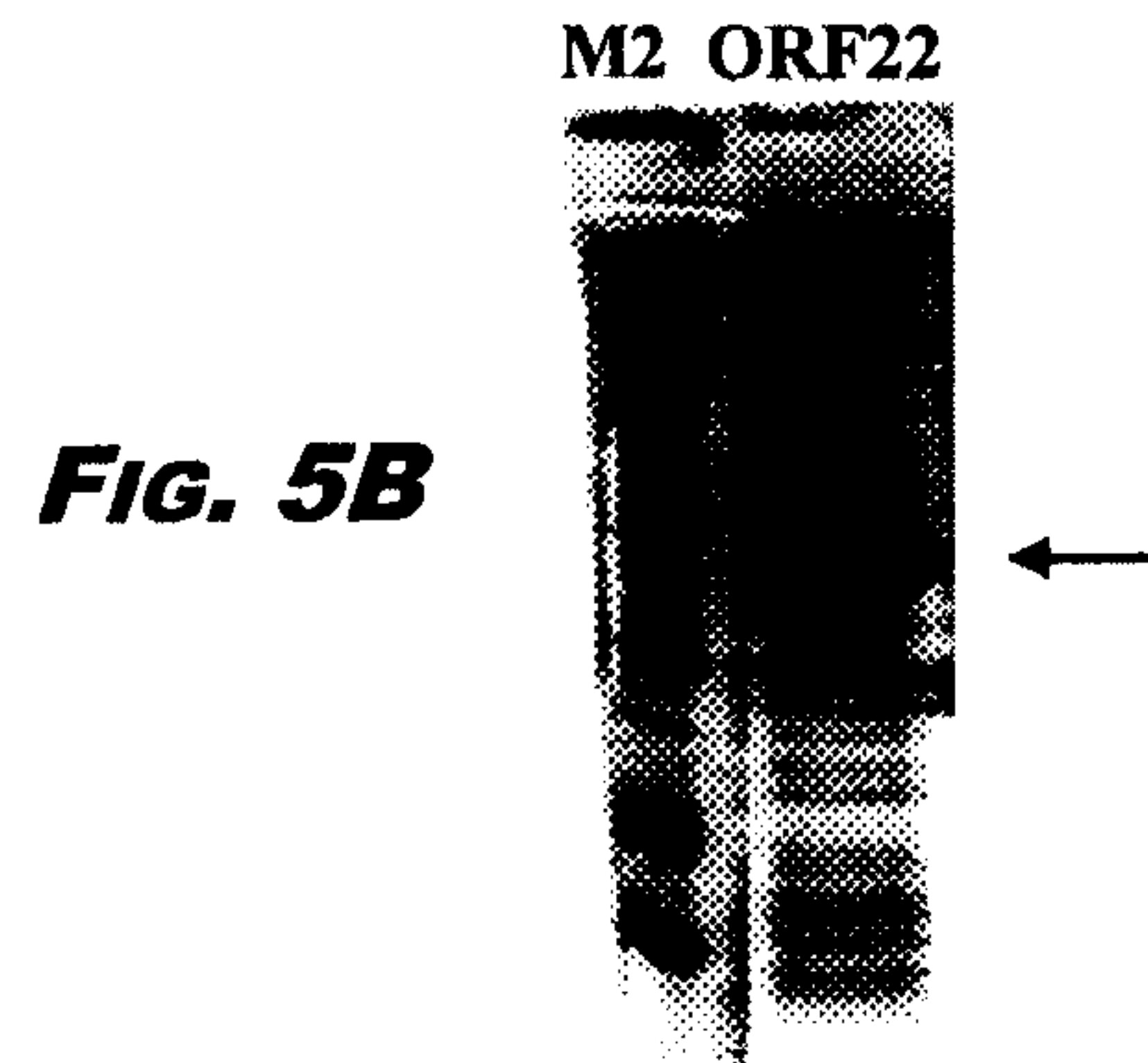
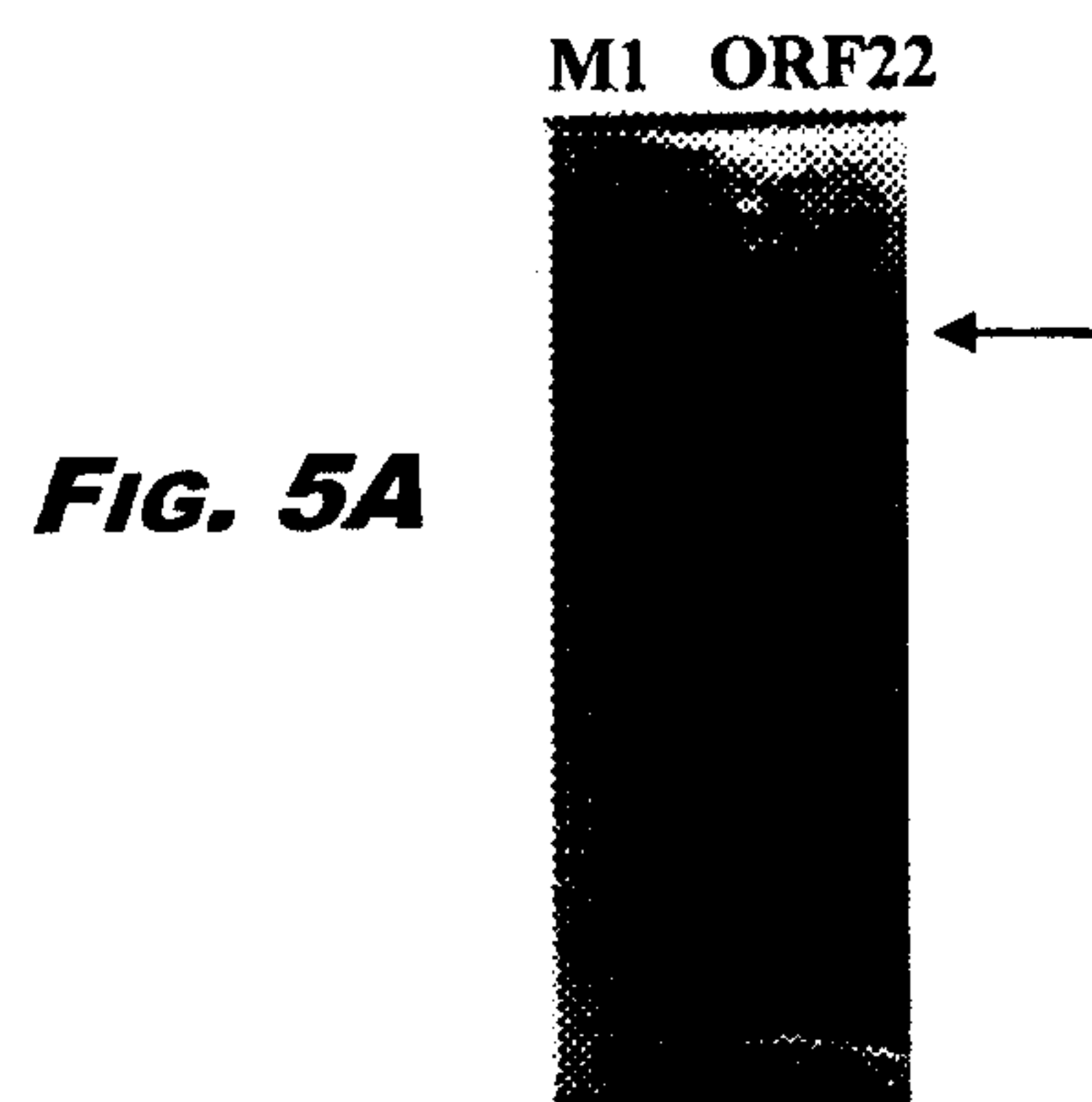
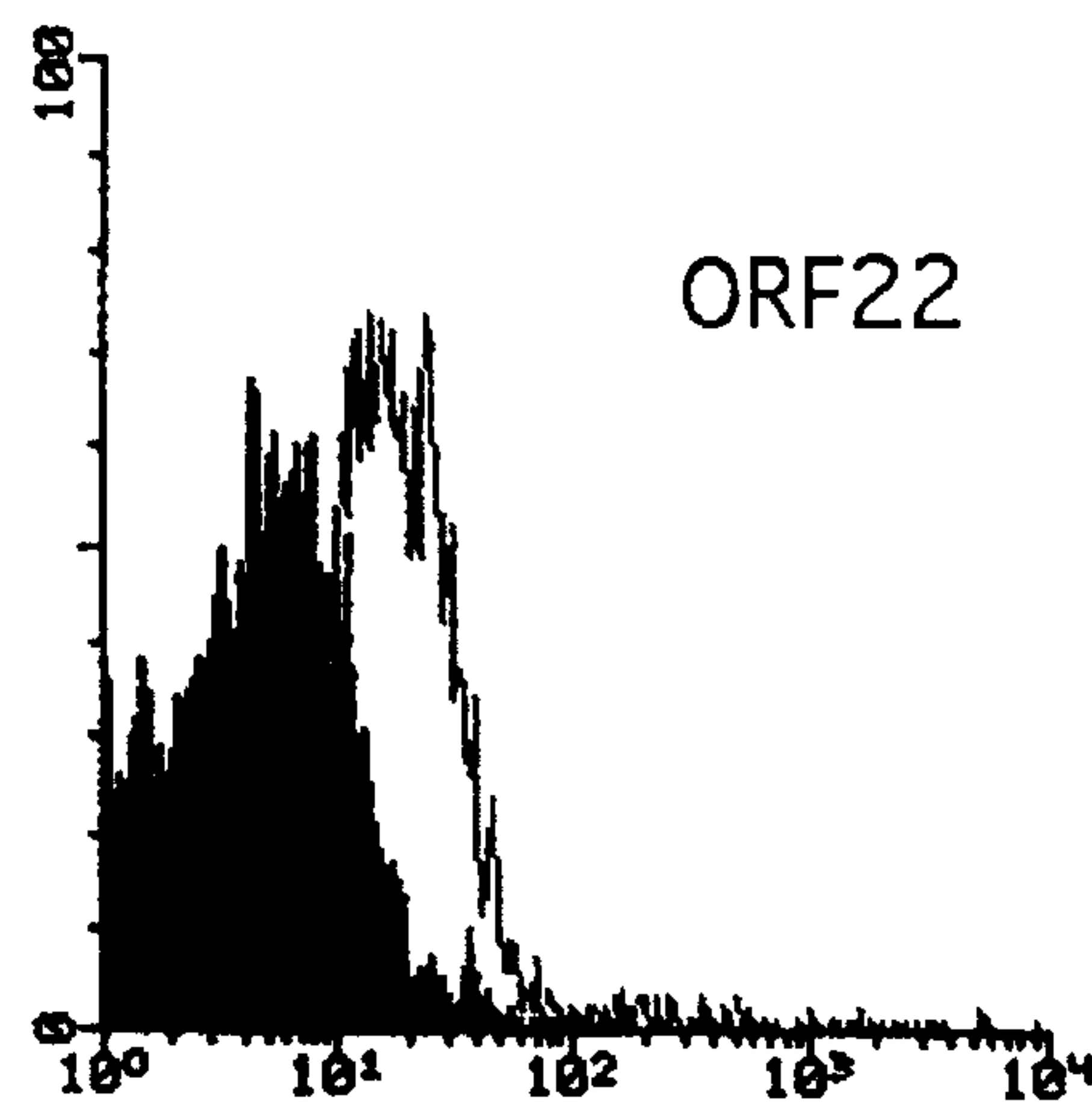
TP OMV



WO 99/24578

PCT/IB98/01665

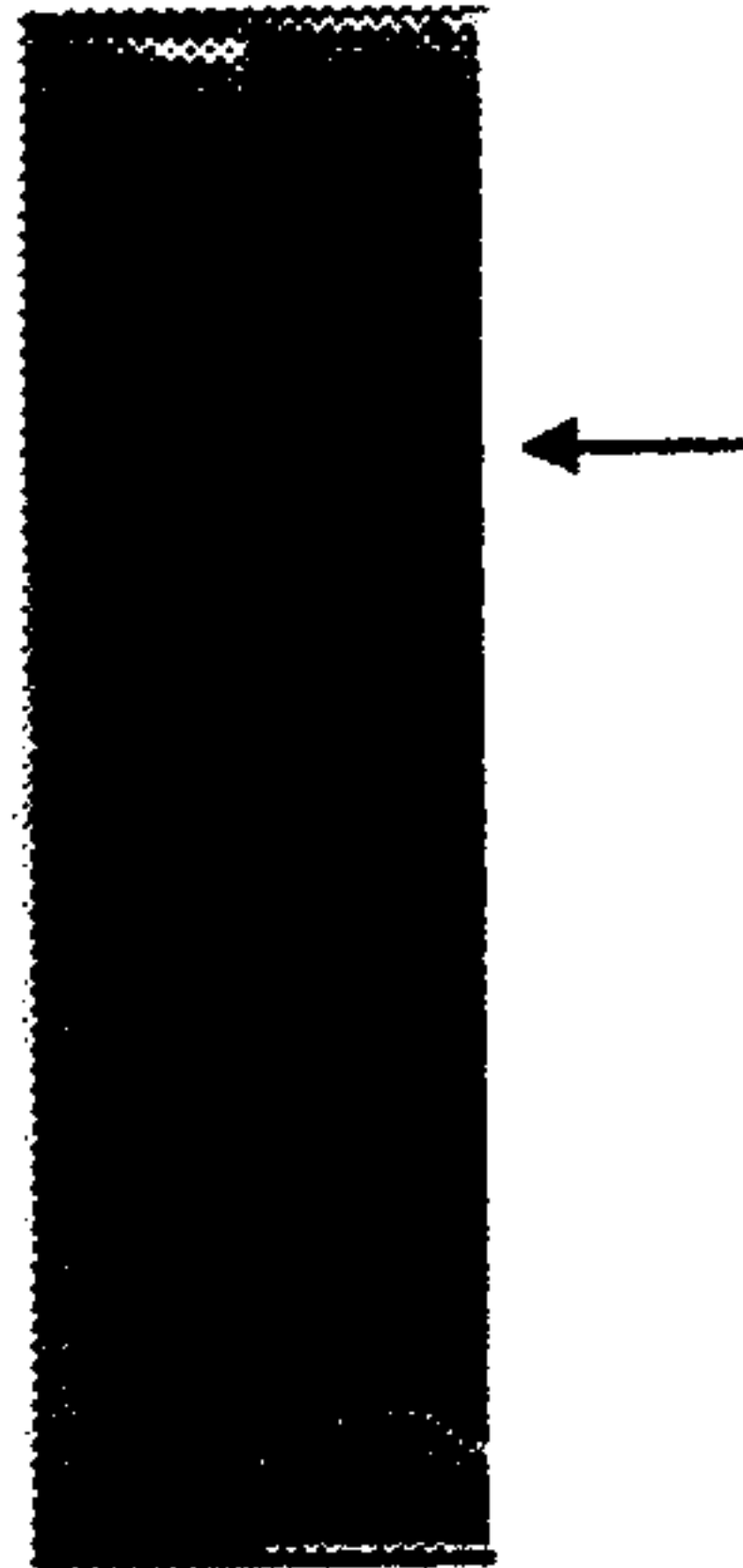
6/24

FIGURE 5**FIG. 5C**

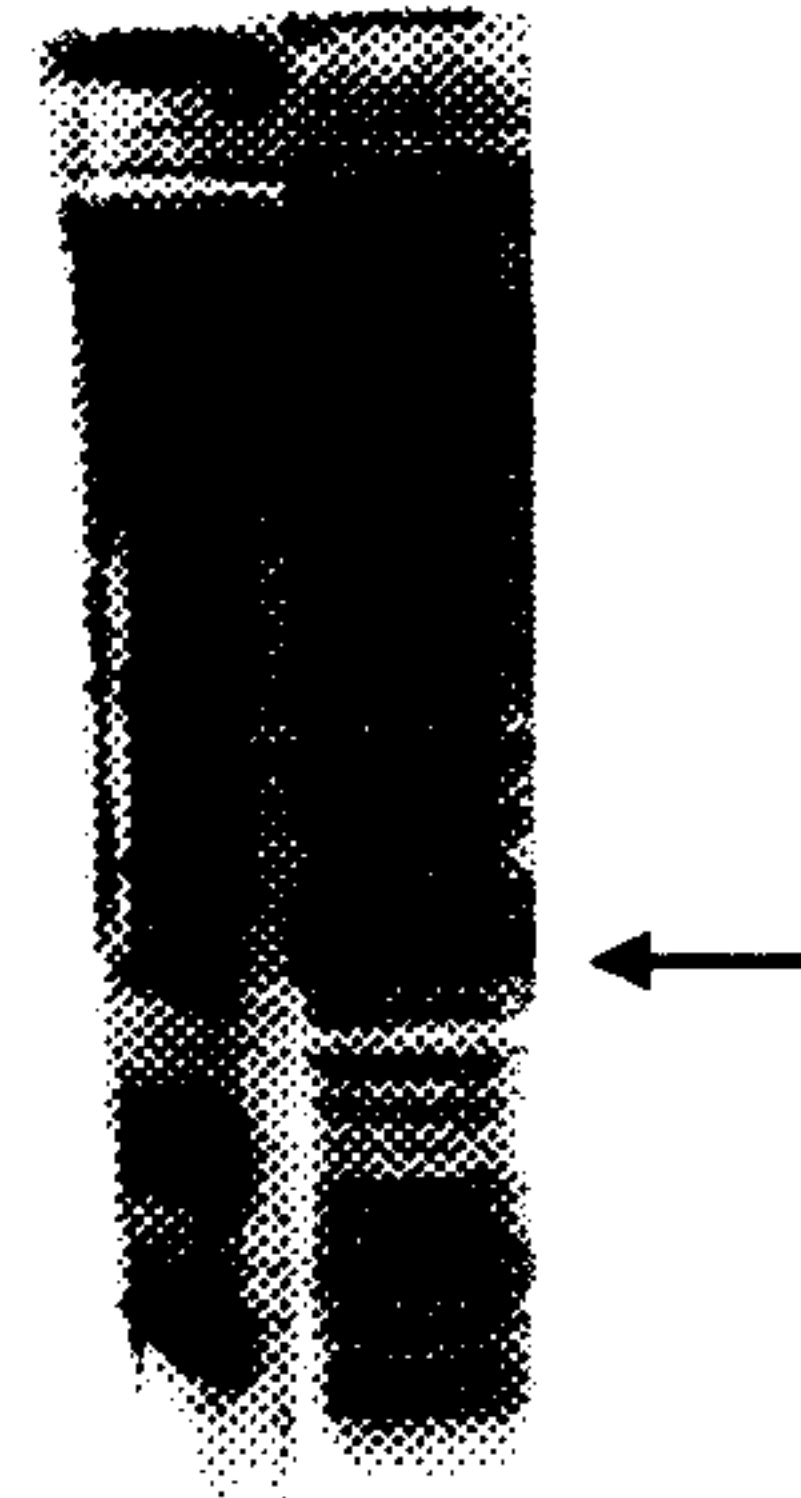
7/24

FIGURE 6**Fig. 6A**

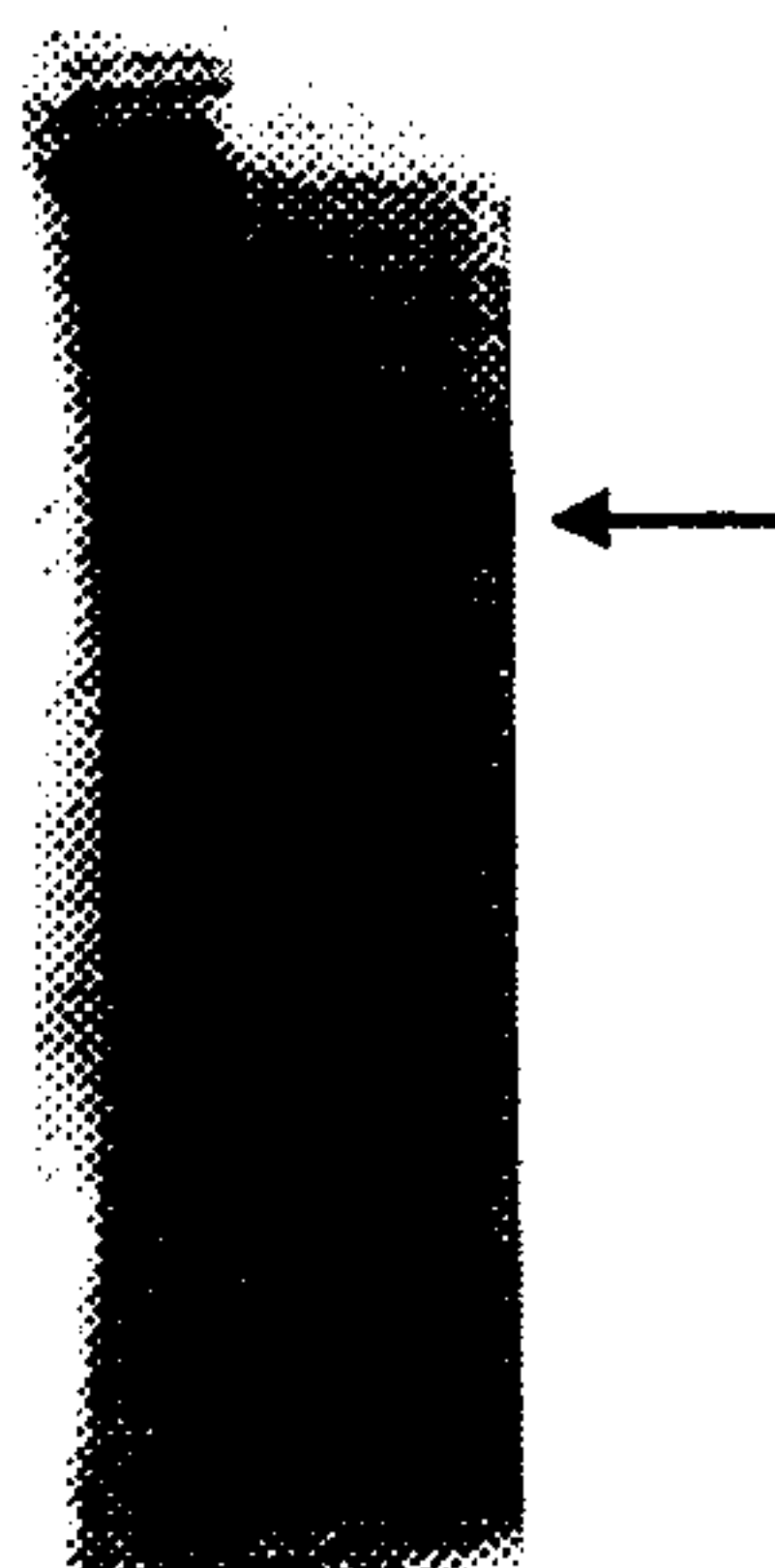
M1 ORF28

**Fig. 6B**

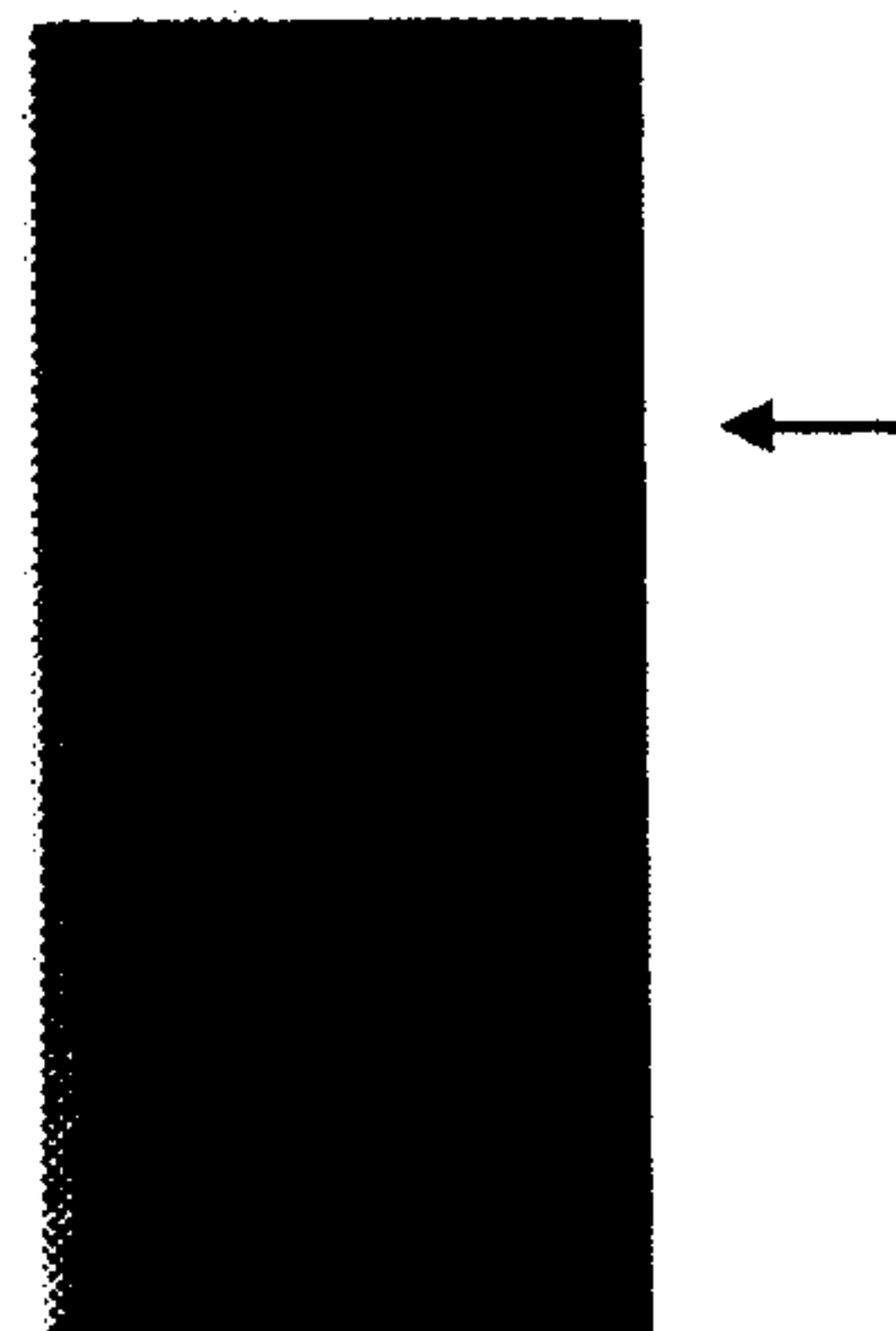
M2 ORF28

**FIGURE 7****Fig. 7A**

M1 ORF32

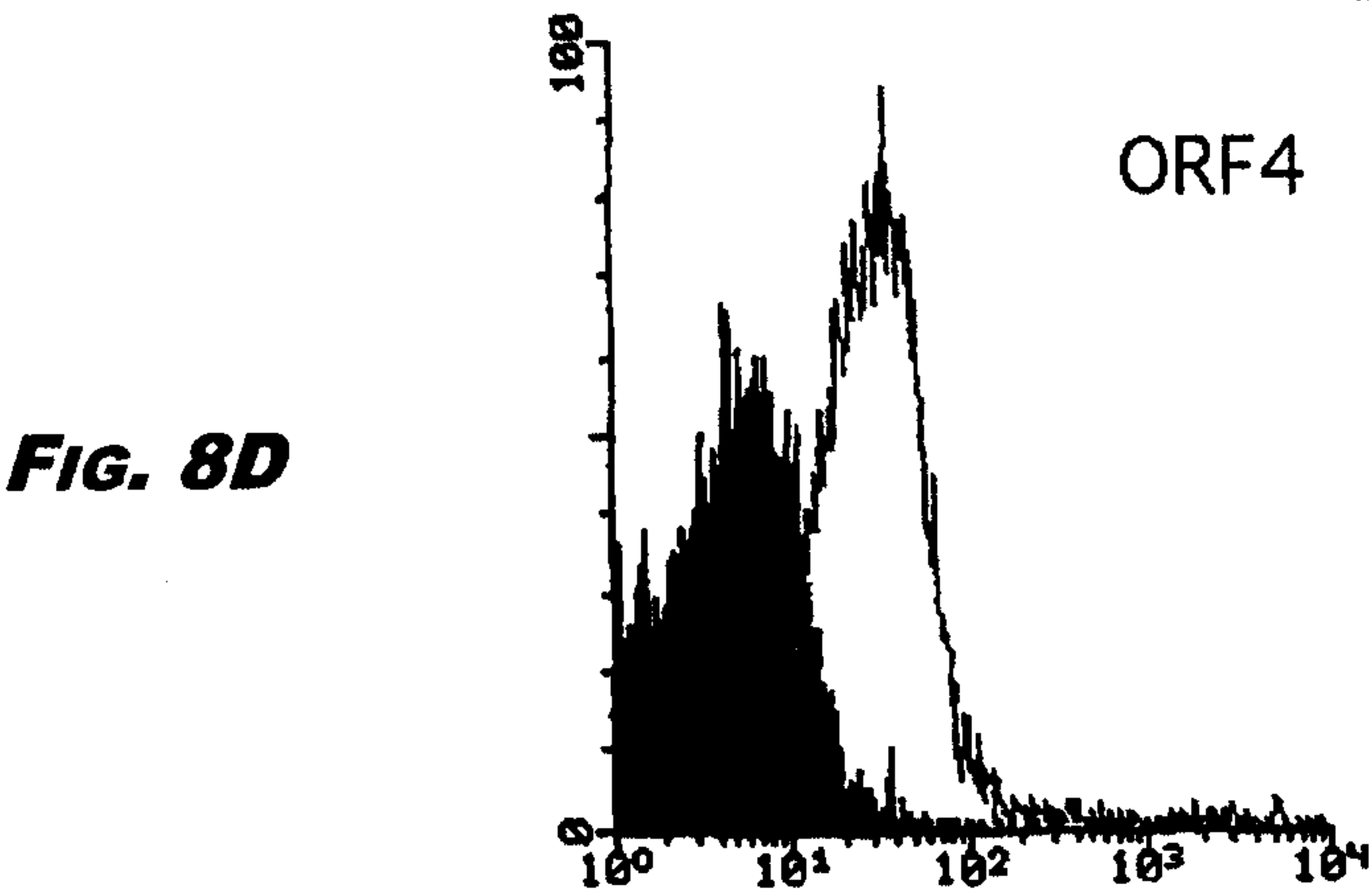
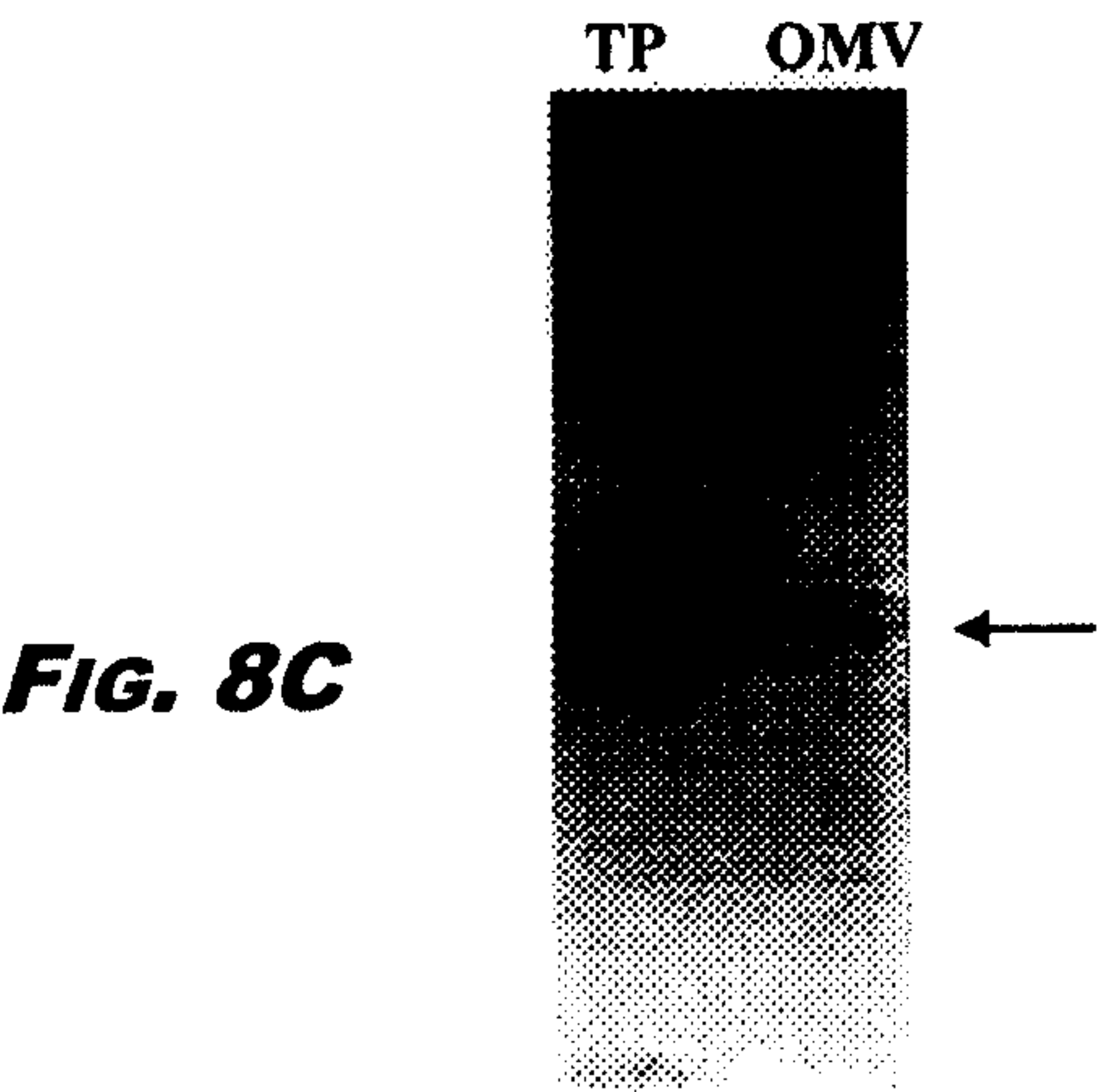
**Fig. 7B**

M1 ORF32



8/24

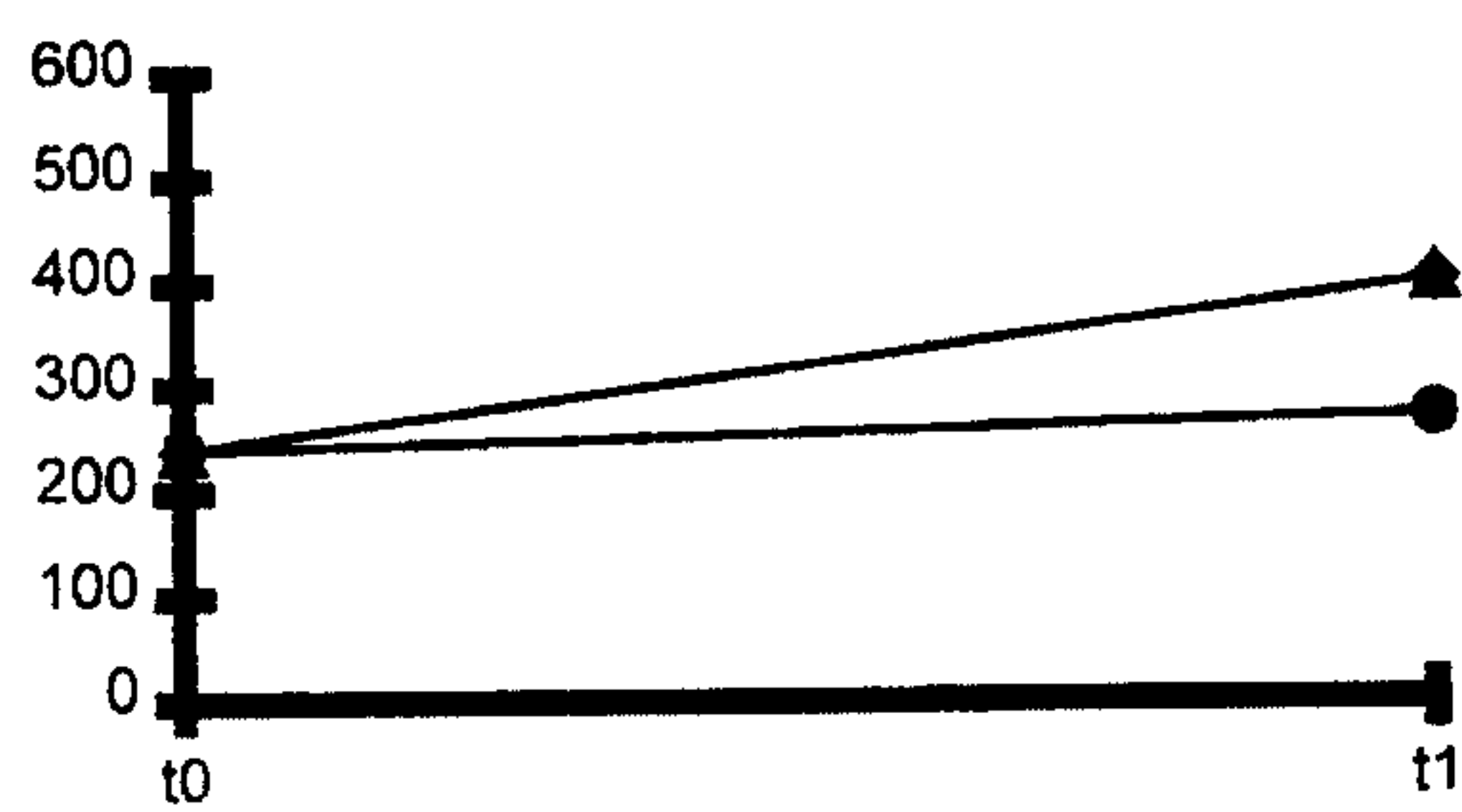
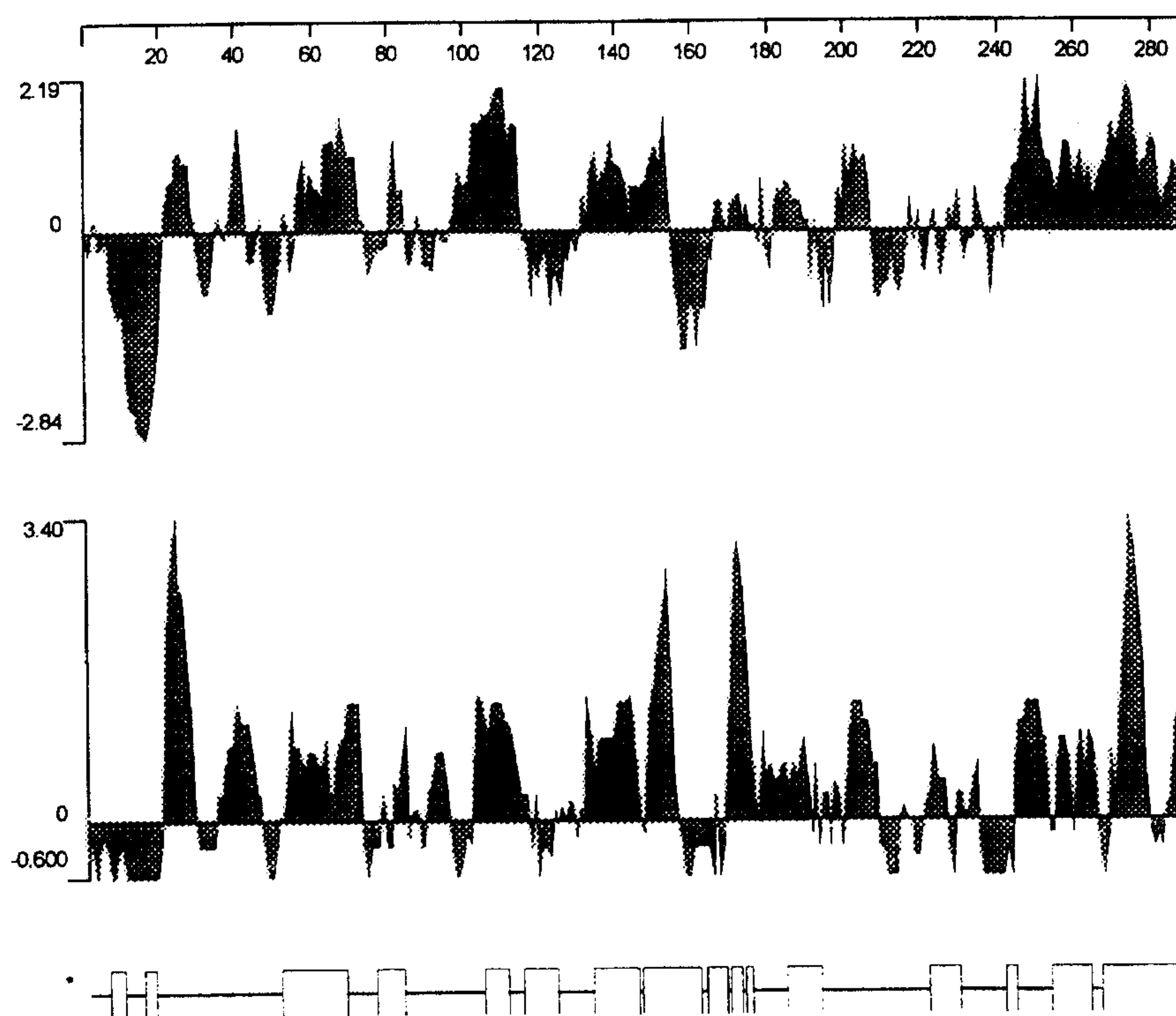
FIGURE 8



WO 99/24578

PCT/IB98/01665

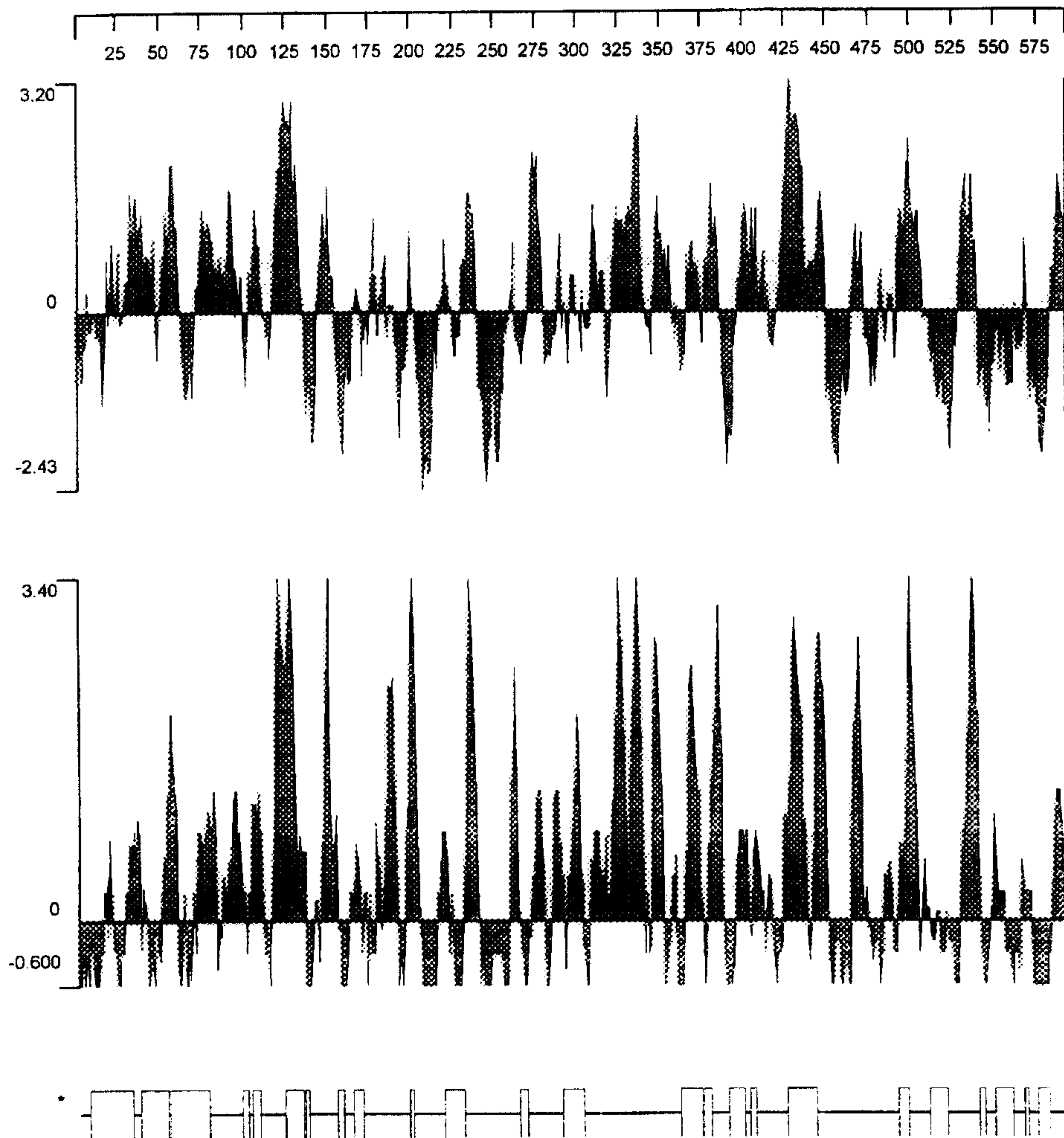
9/24

FIG. 8E**FIG. 8F**

WO 99/24578

PCT/IB98/01665

10/24

FIGURE 9

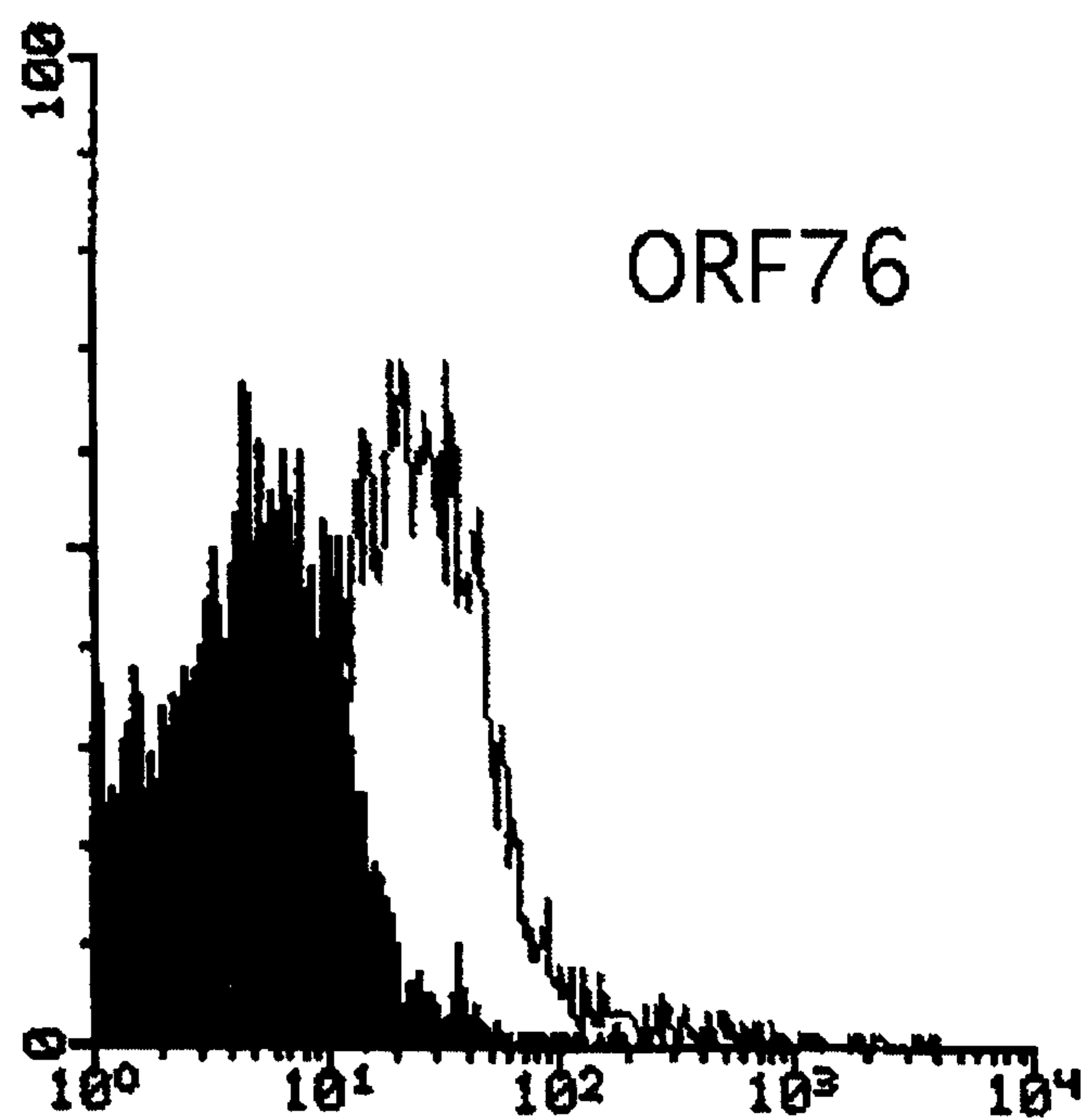
11/24

FIGURE 10**Fig. 10A**

M1 ORF76

**Fig. 10B**

TP OMV

**Fig. 10C**

WO 99/24578

PCT/IB98/01665

12/24

FIGURE 11

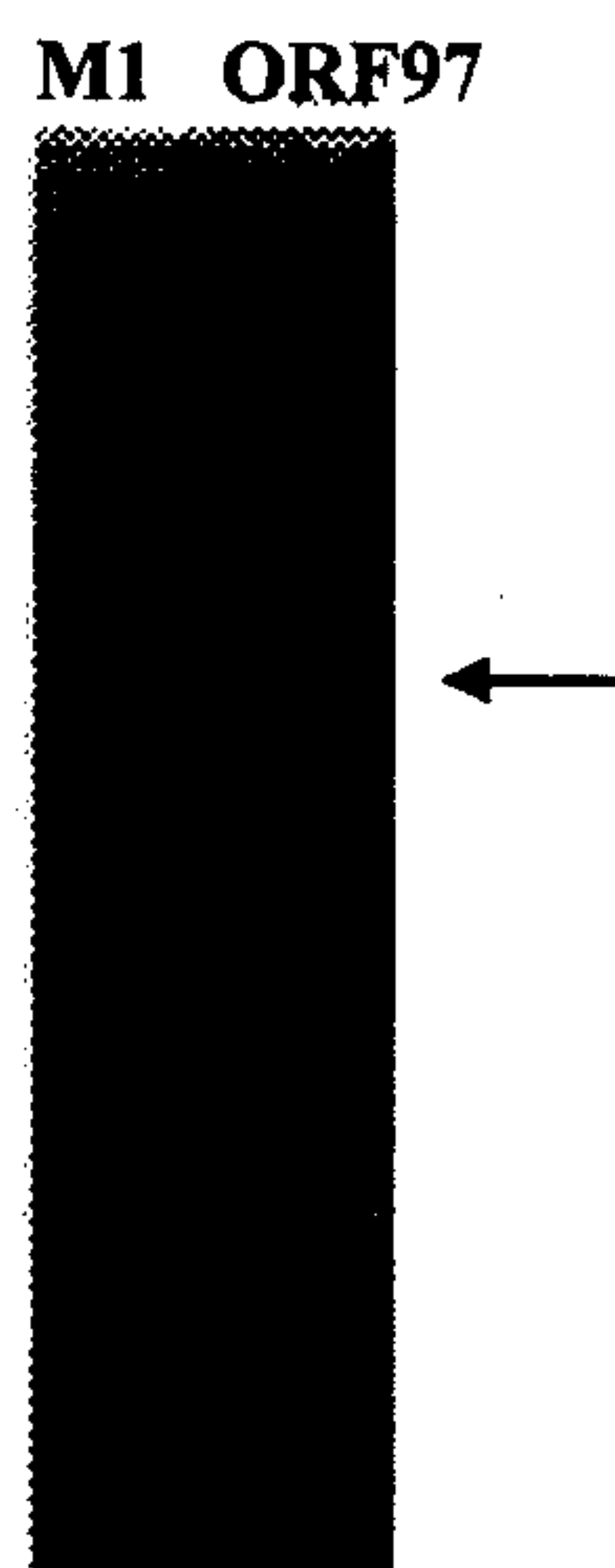
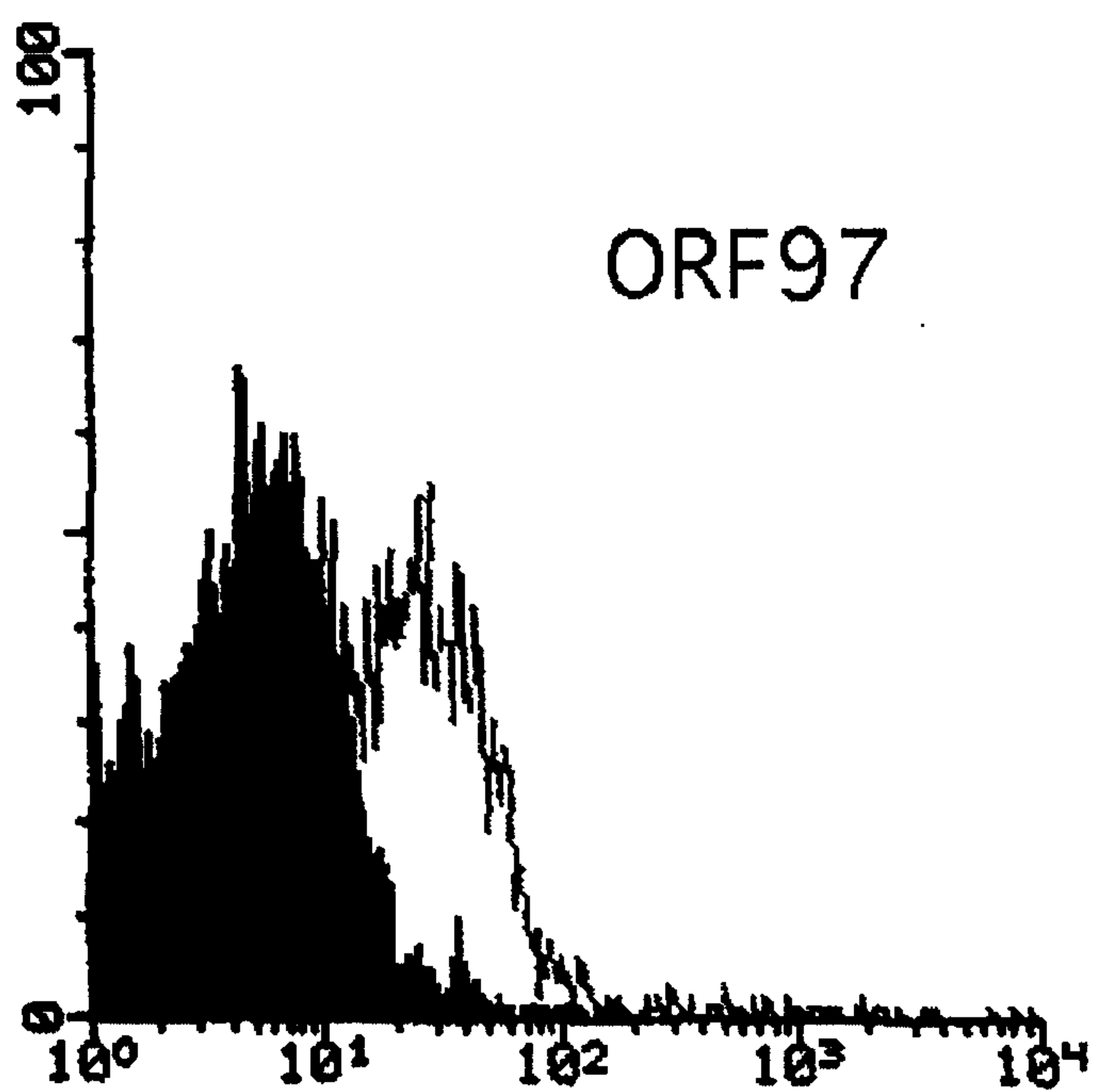
M1 ORF89



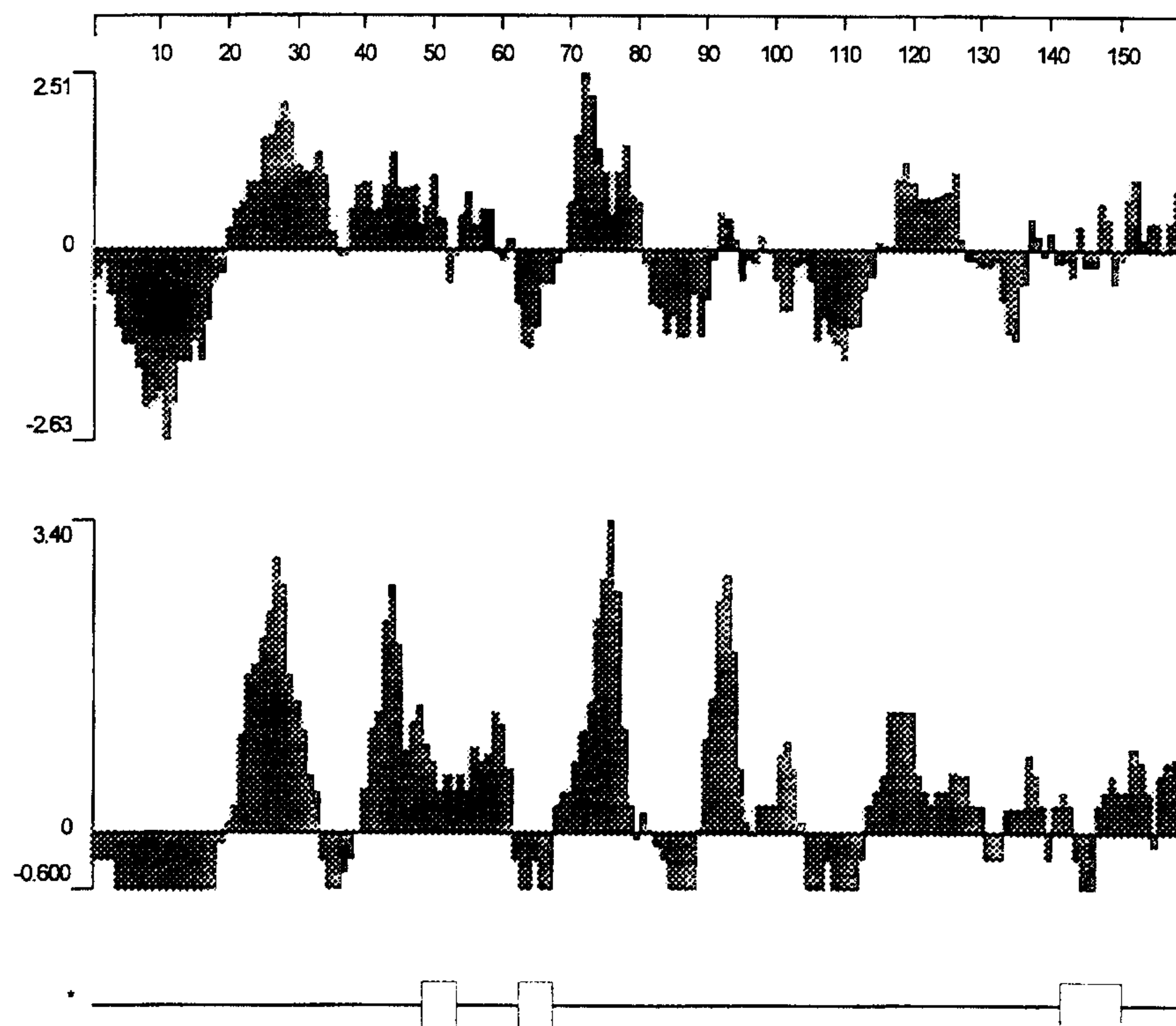
WO 99/24578

PCT/IB98/01665

13/24

FIGURE 12**Fig. 12A****Fig. 12B****Fig. 12C****Fig. 12D**

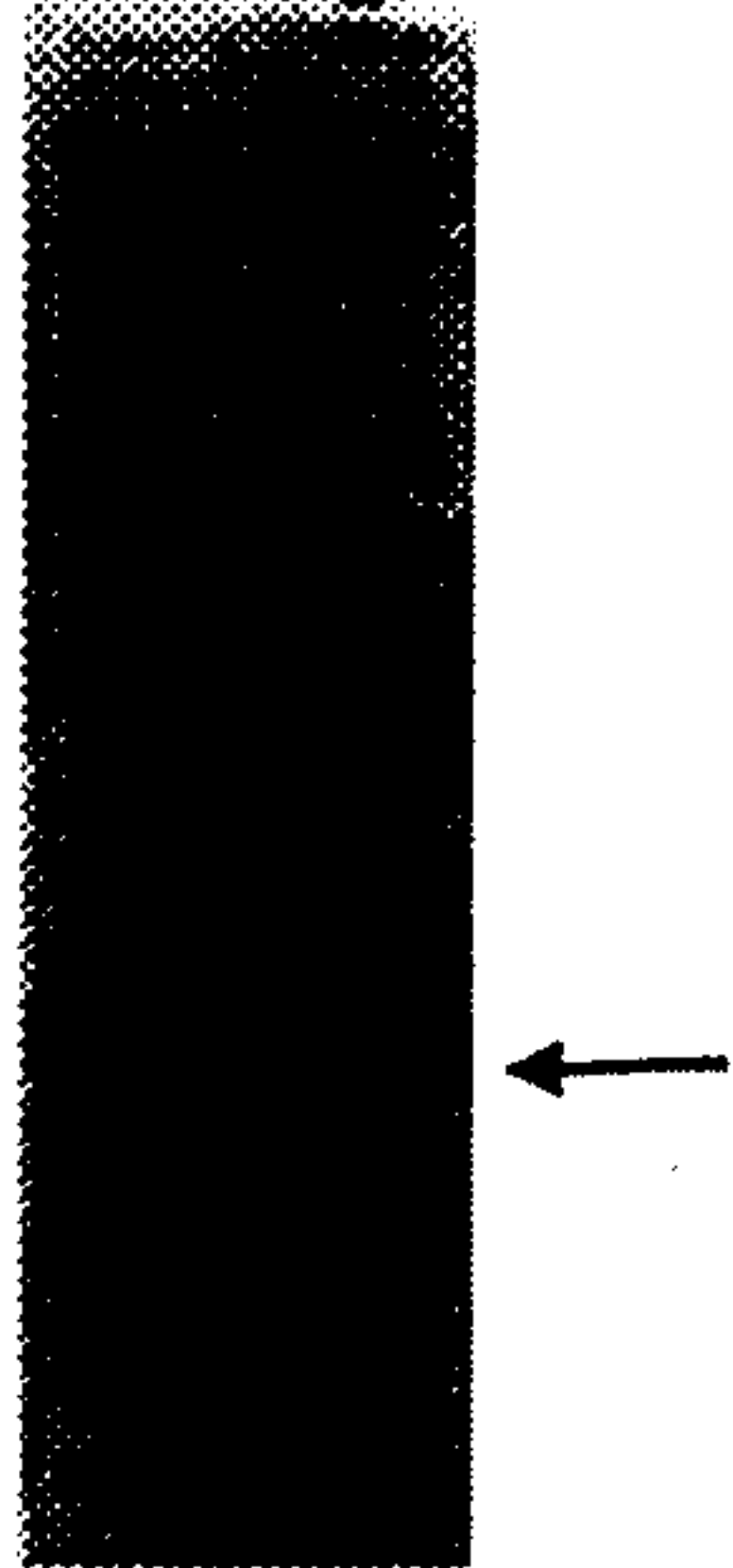
14/24

FIG. 12E

15/24

FIGURE 13**Fig. 13A**

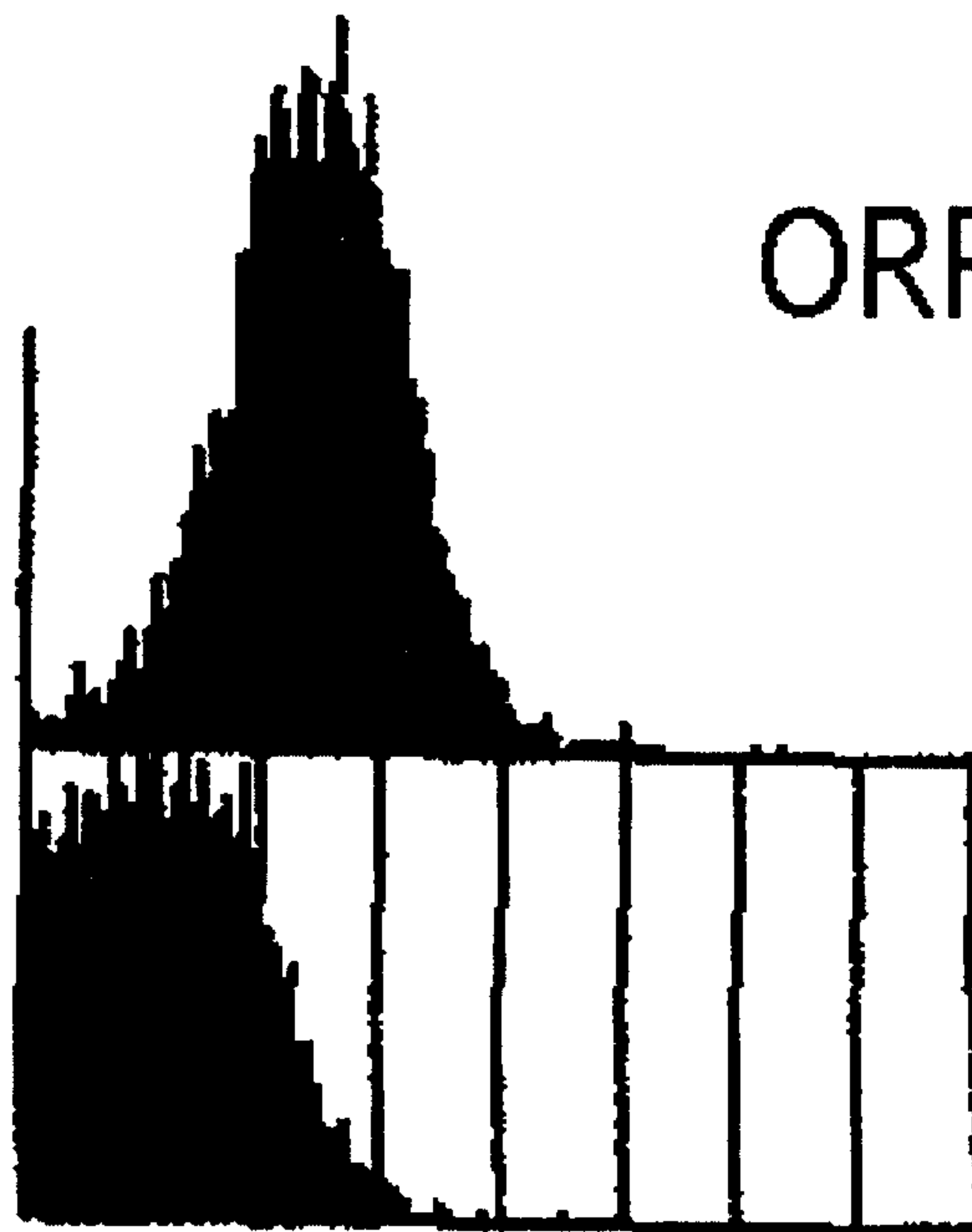
M1 ORF106

**Fig. 13B**

M2 ORF106

**Fig. 13C**

ORF 106



WO 99/24578

PCT/IB98/01665

16/24

FIGURE 14

FIG. 14A

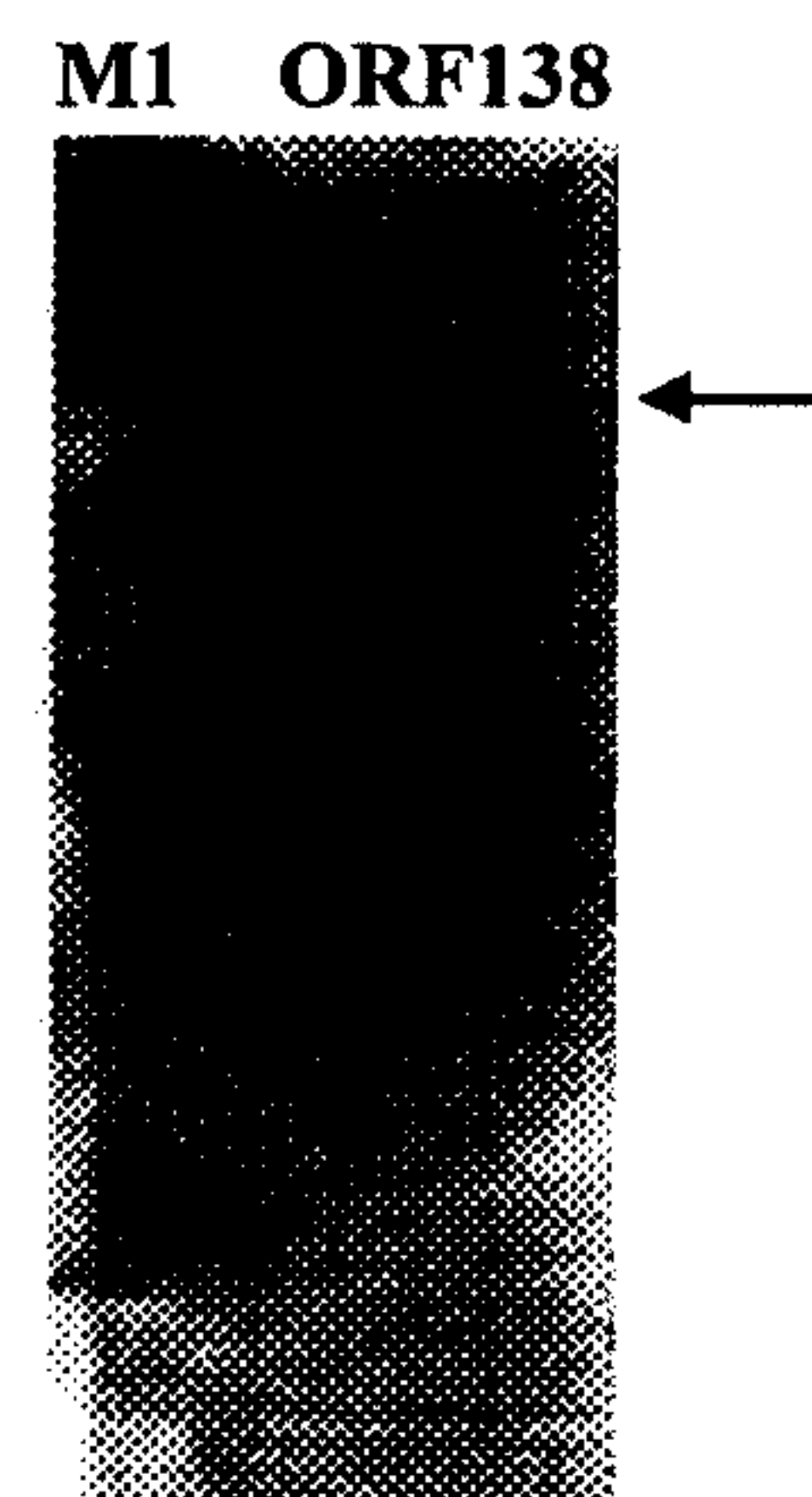
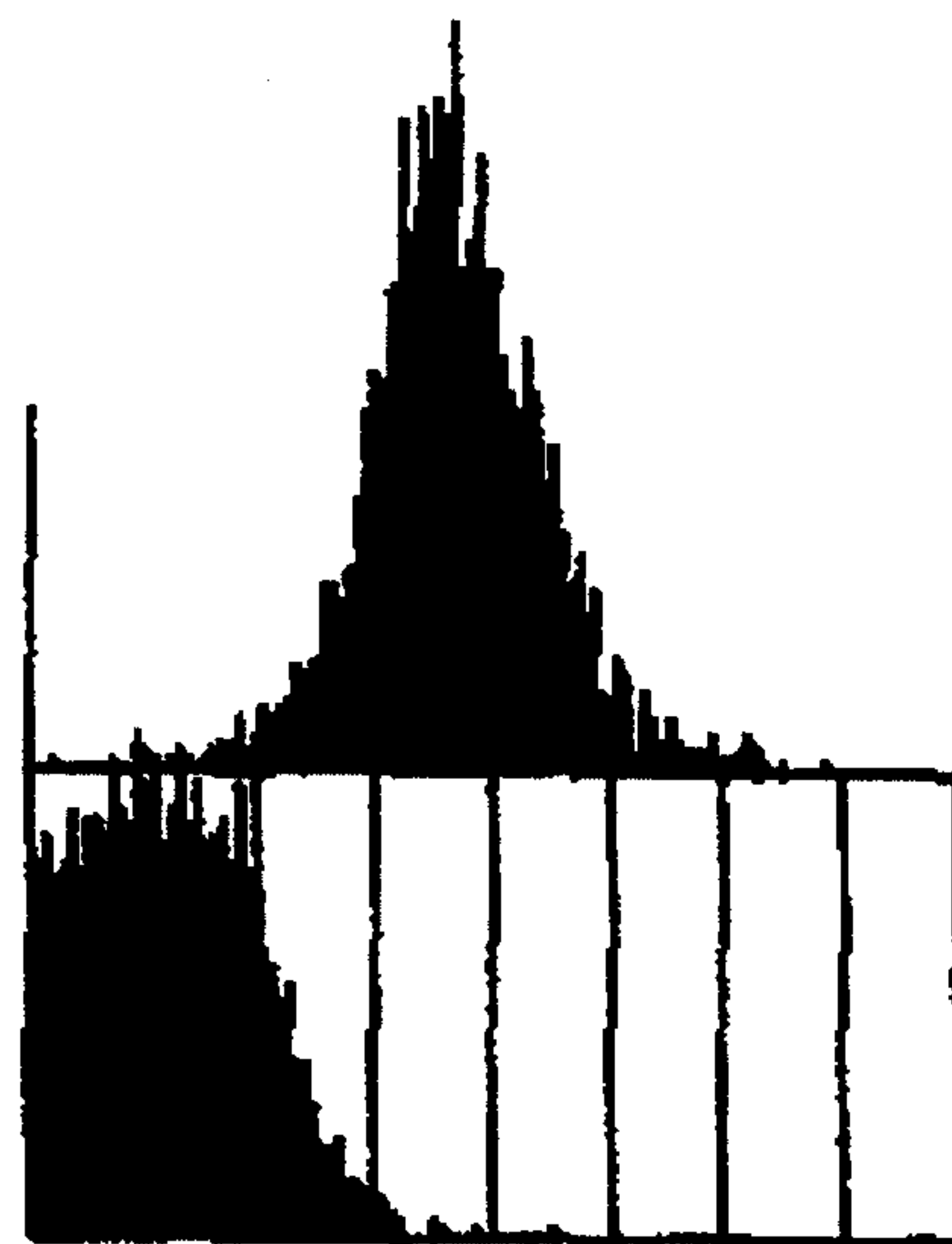


FIG. 14B



17/24

FIGURE 15

Fig. 15A

M1 ORF23



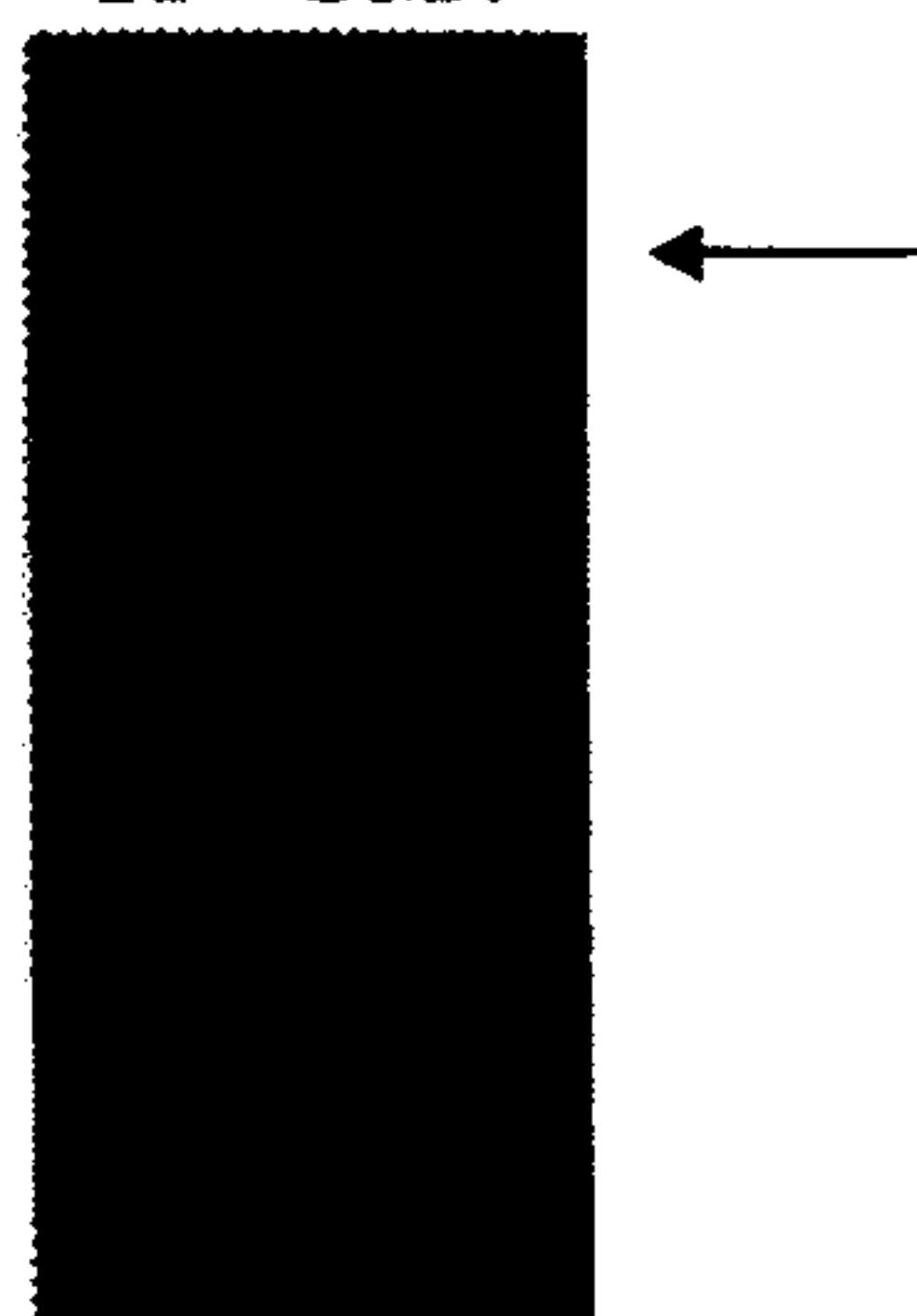
Fig. 15B

M2 ORF23



Fig 15C

TP OMV



WO 99/24578

PCT/IB98/01665

18/24

FIGURE 16

Fig. 16A

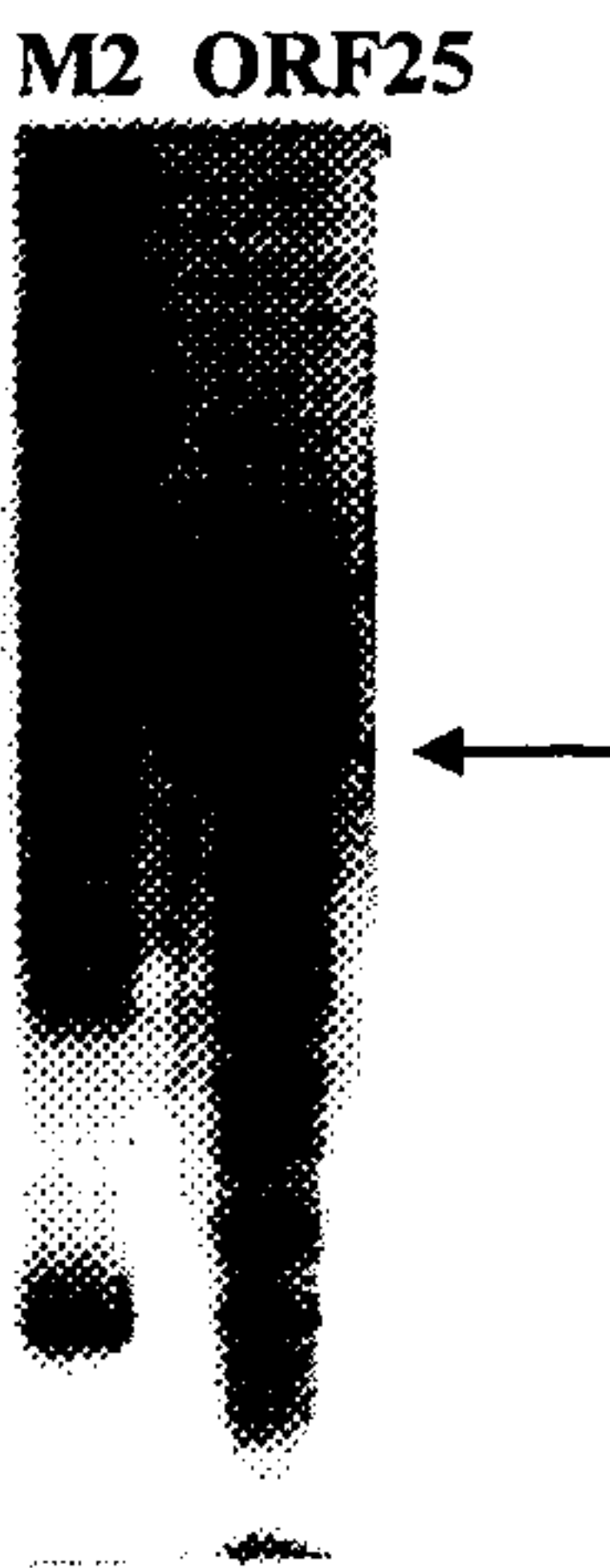


Fig. 16B

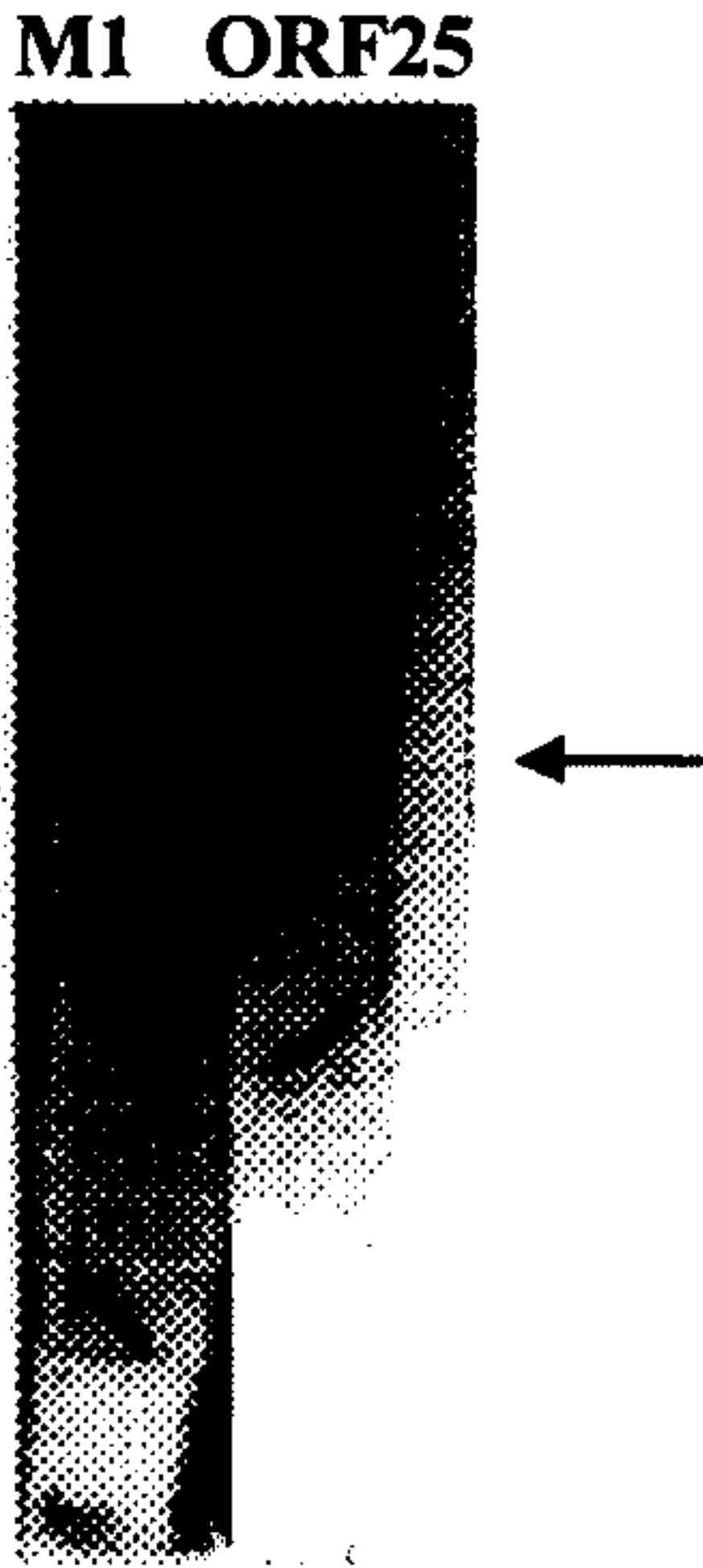


Fig. 16C

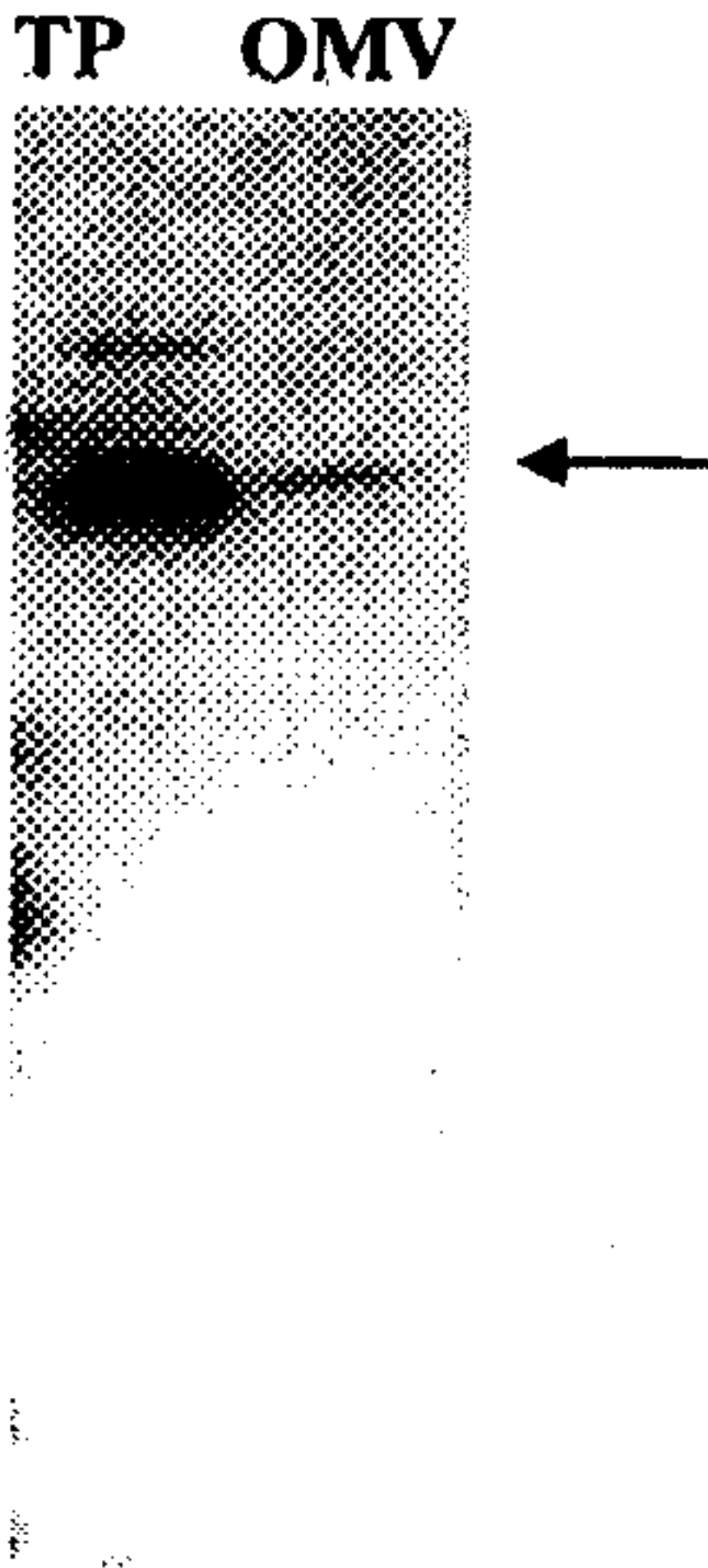
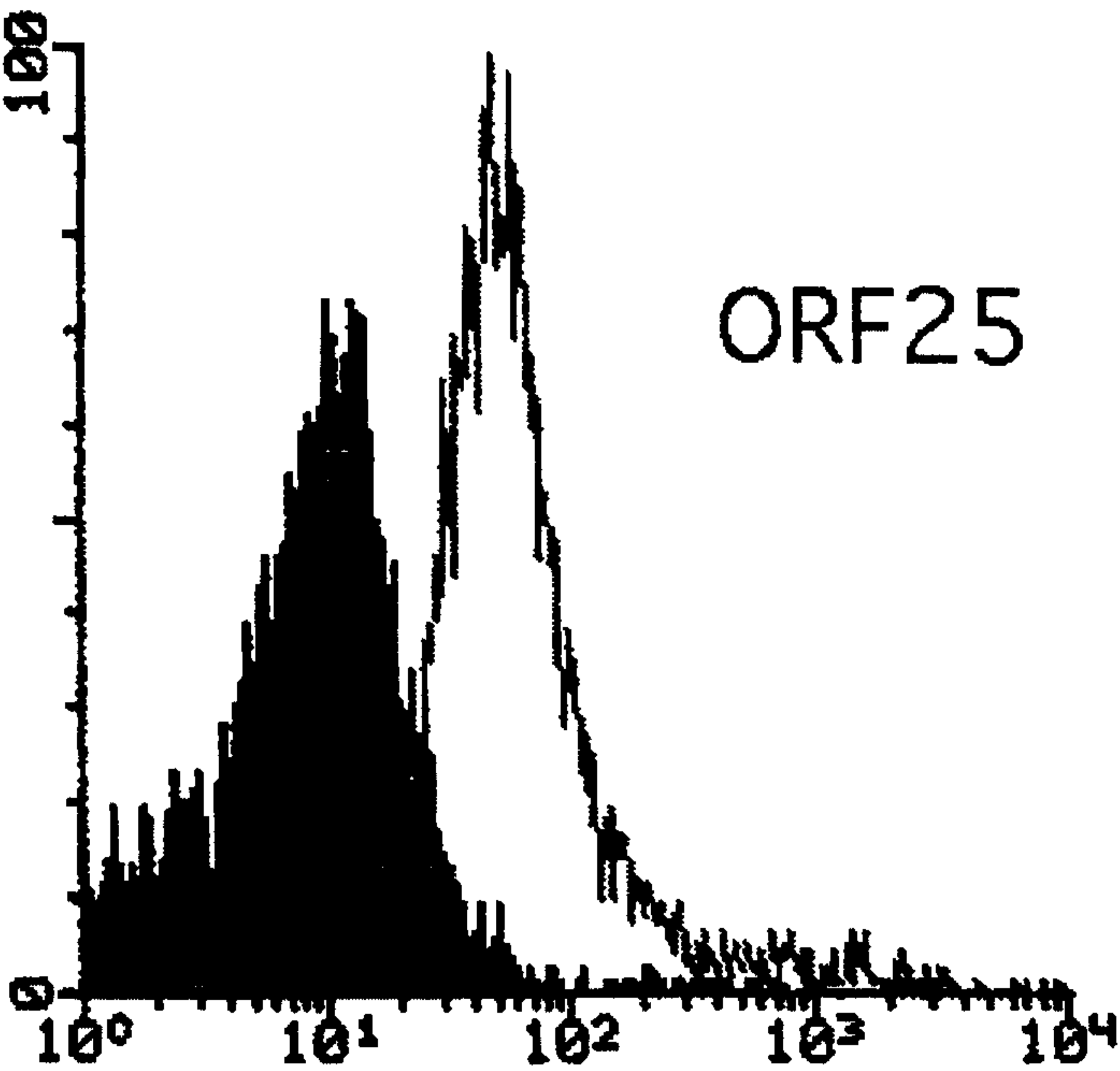


Fig. 16D



19/24

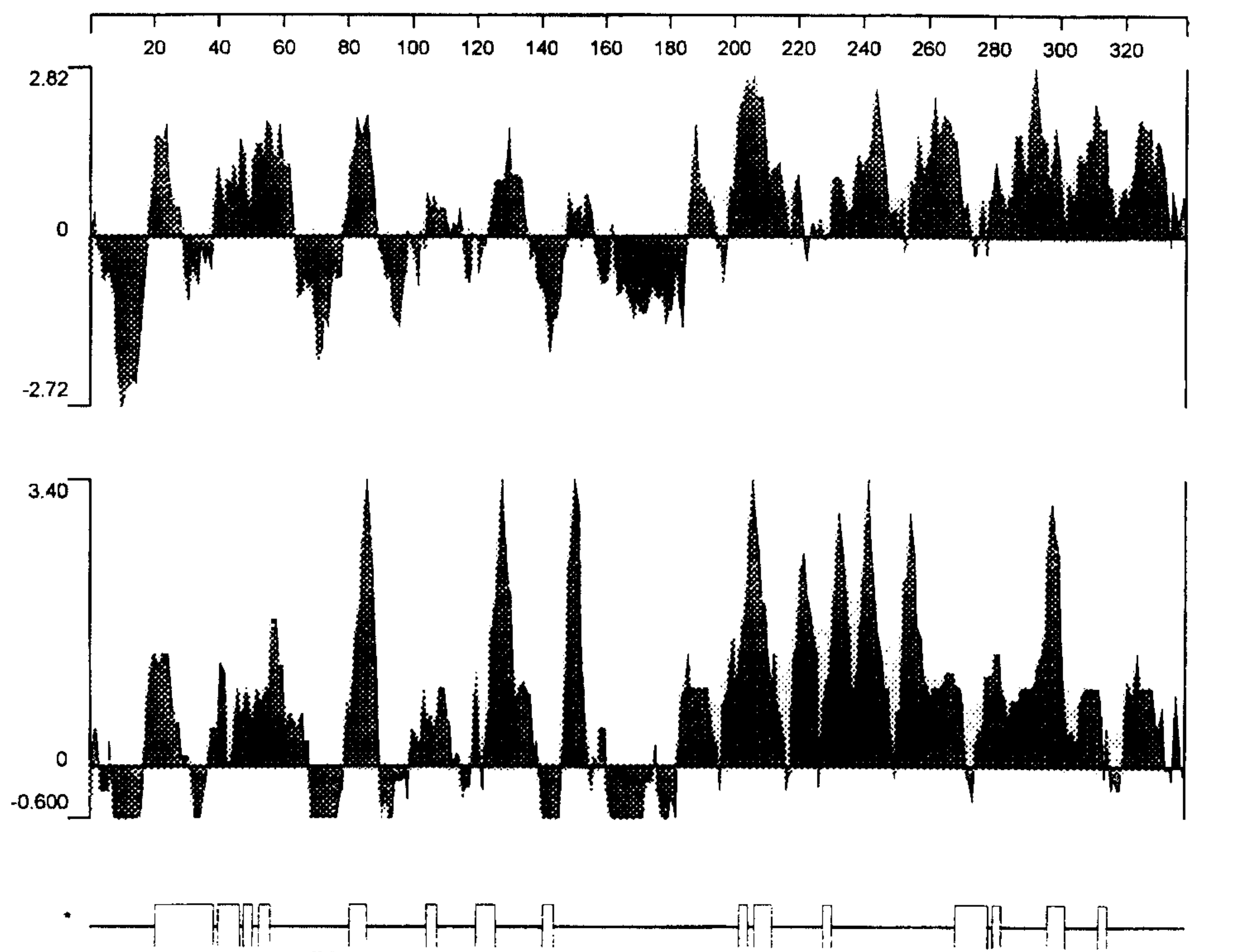
FIG. 16E

FIGURE 17

Fig. 17A

M1 ORF27



Fig. 17B

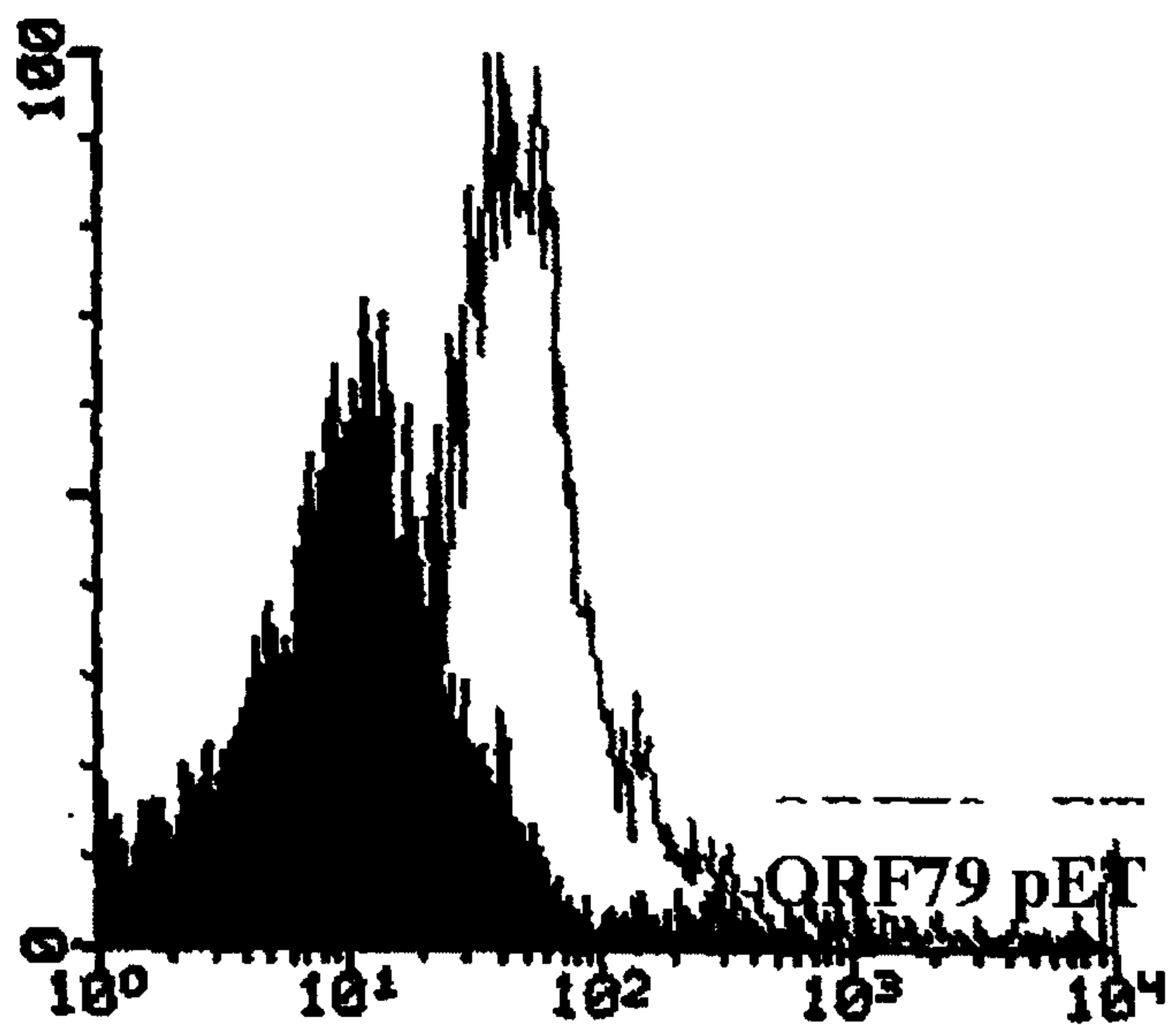
M2 ORF27



21/24

FIGURE 18**Fig. 18A**

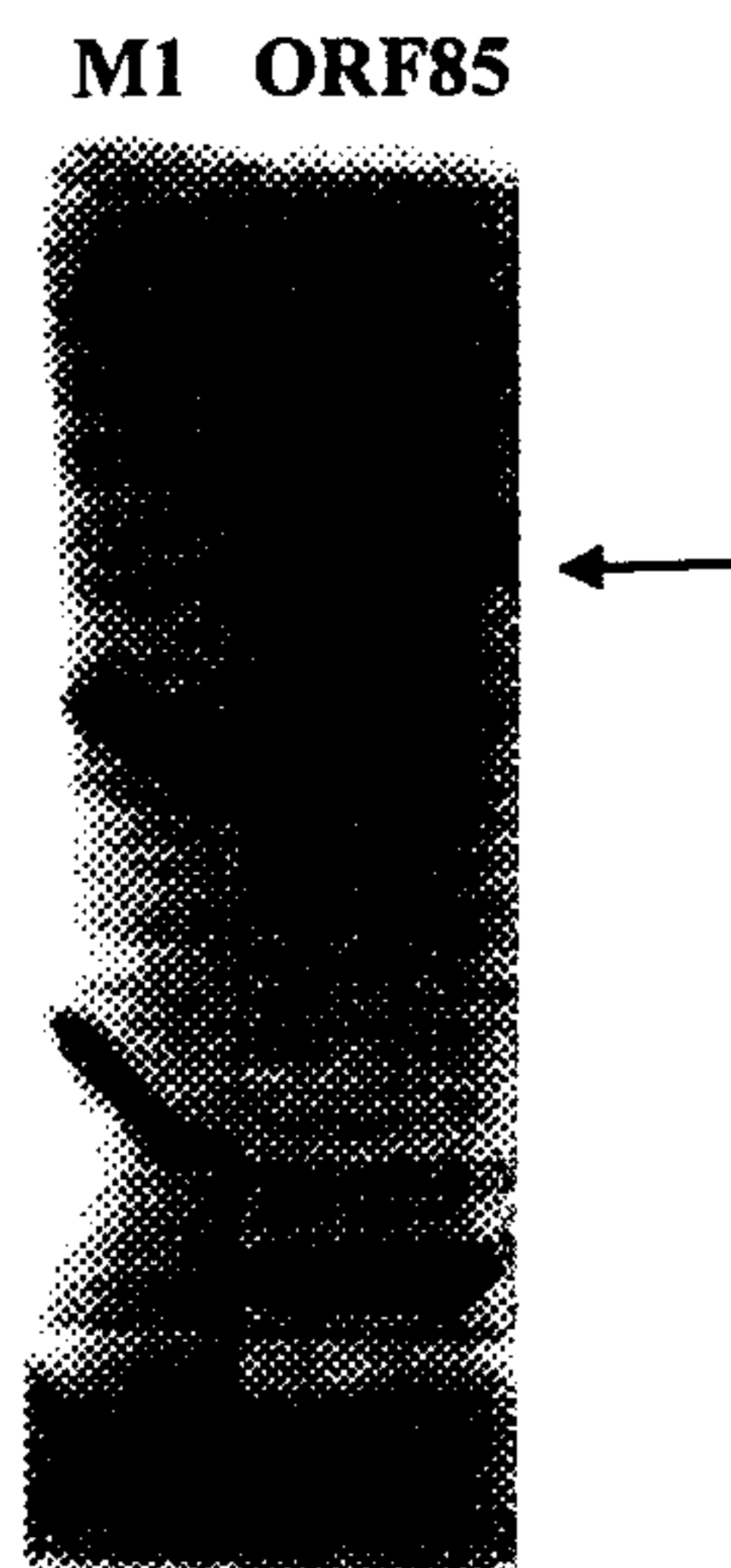
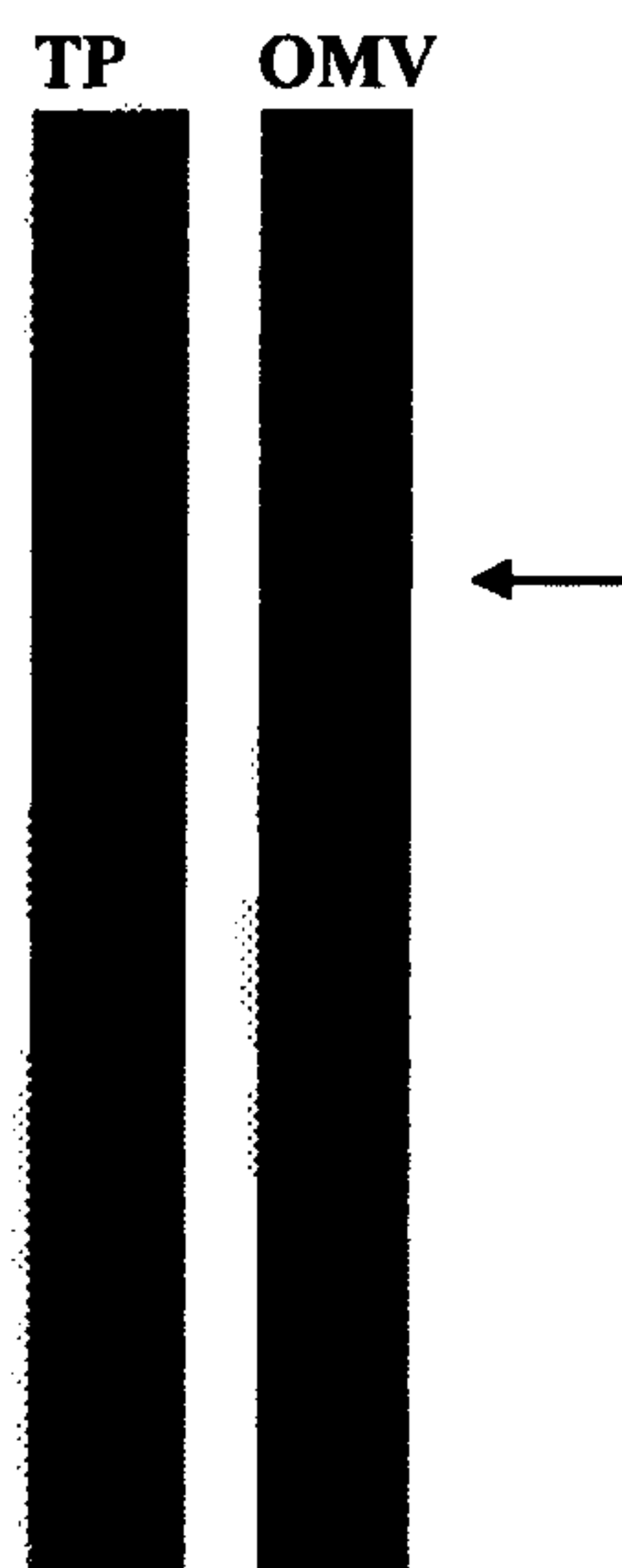
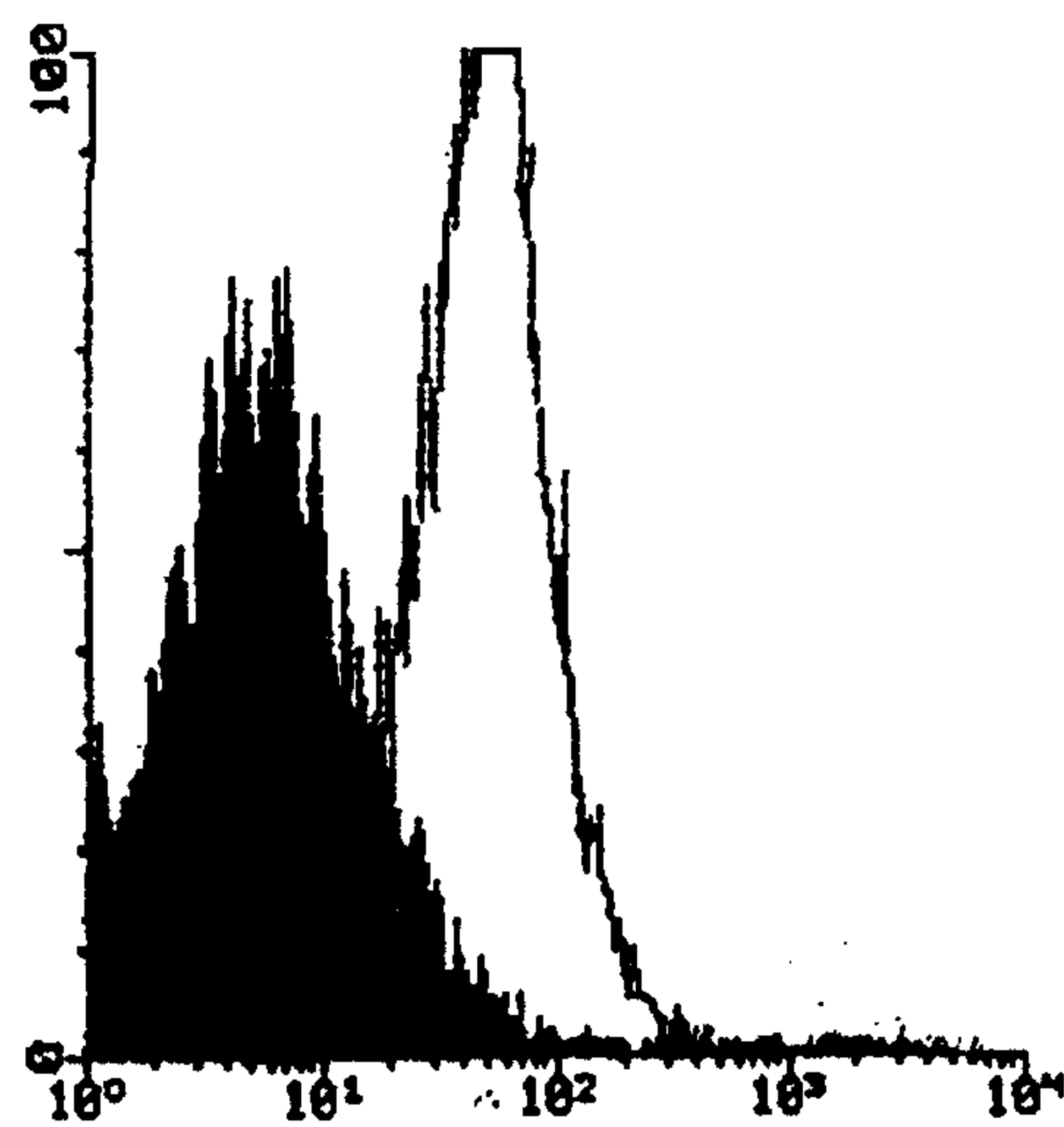
M1 ORF79

**Fig. 18B**

WO 99/24578

PCT/IB98/01665

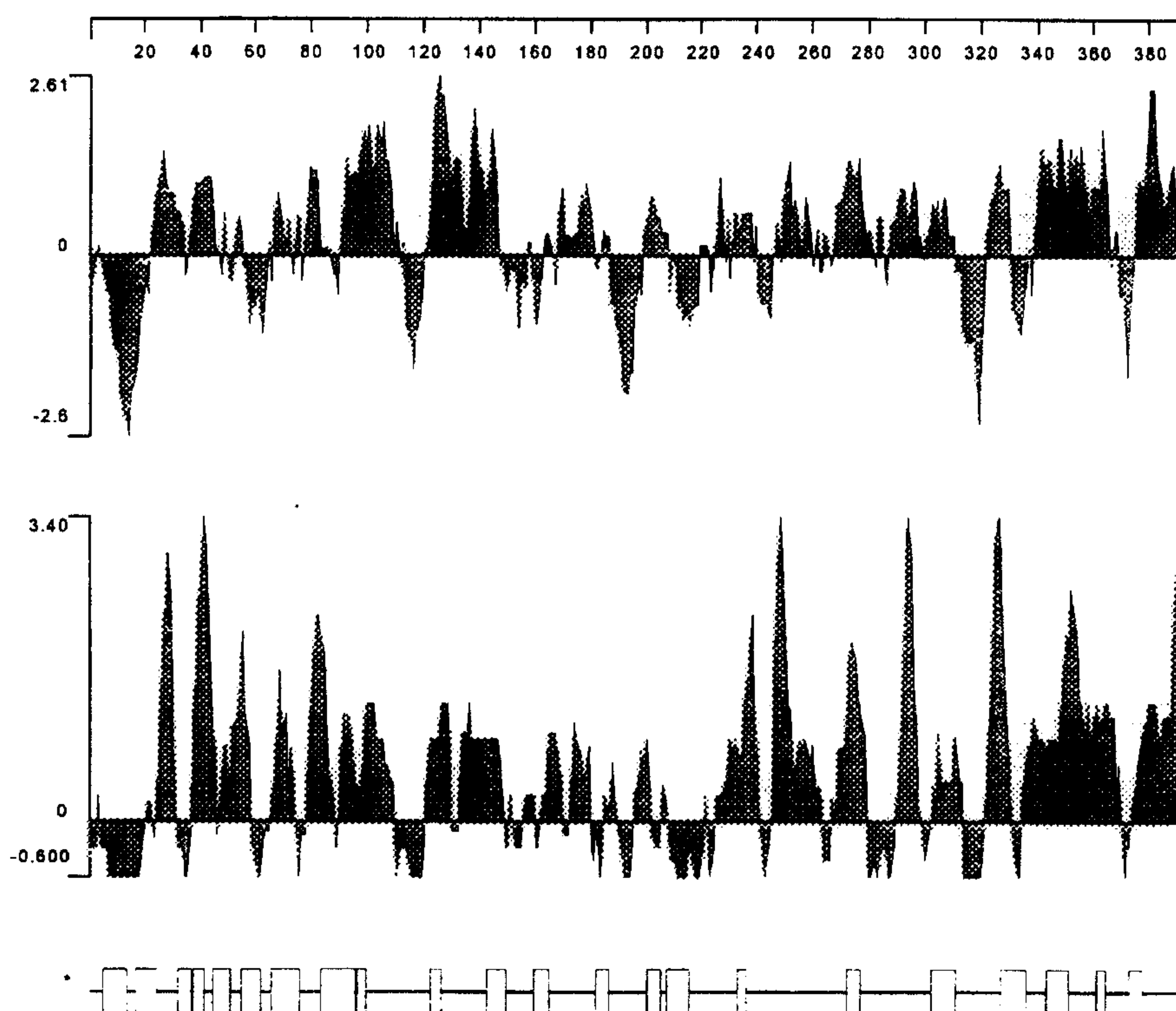
22/24

FIGURE 19**Fig. 19A****Fig. 19B****Fig. 19C**

WO 99/24578

PCT/IB98/01665

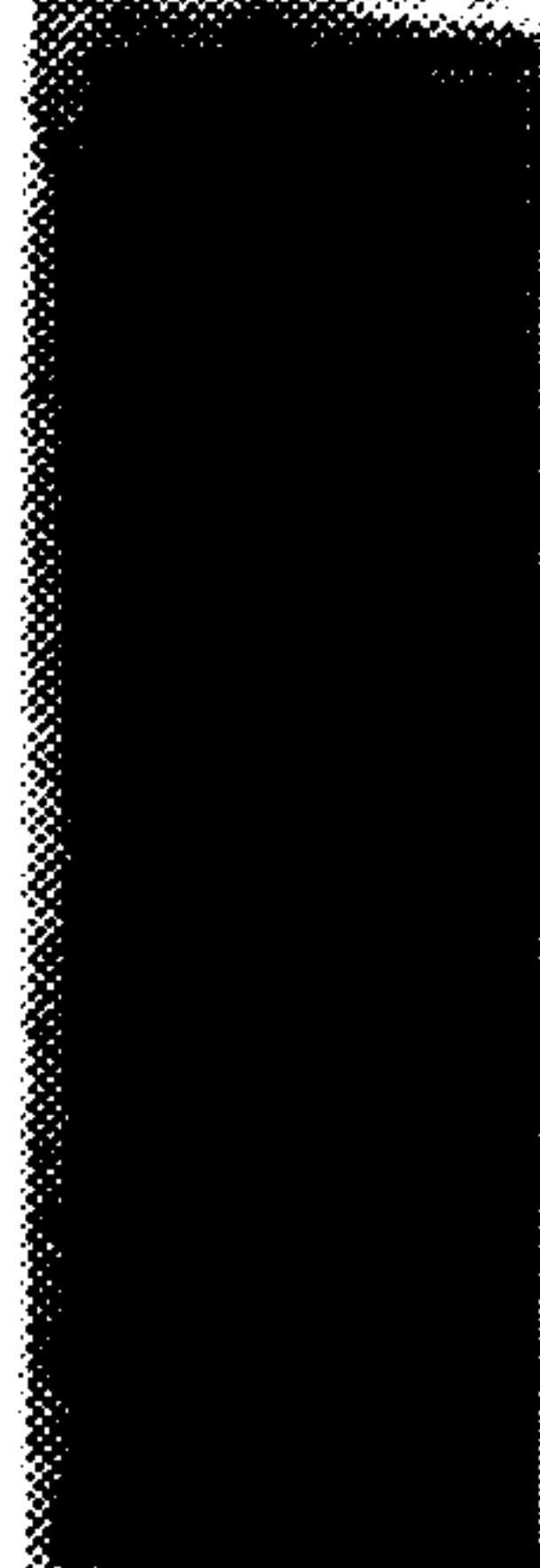
23/24

Fig 19D

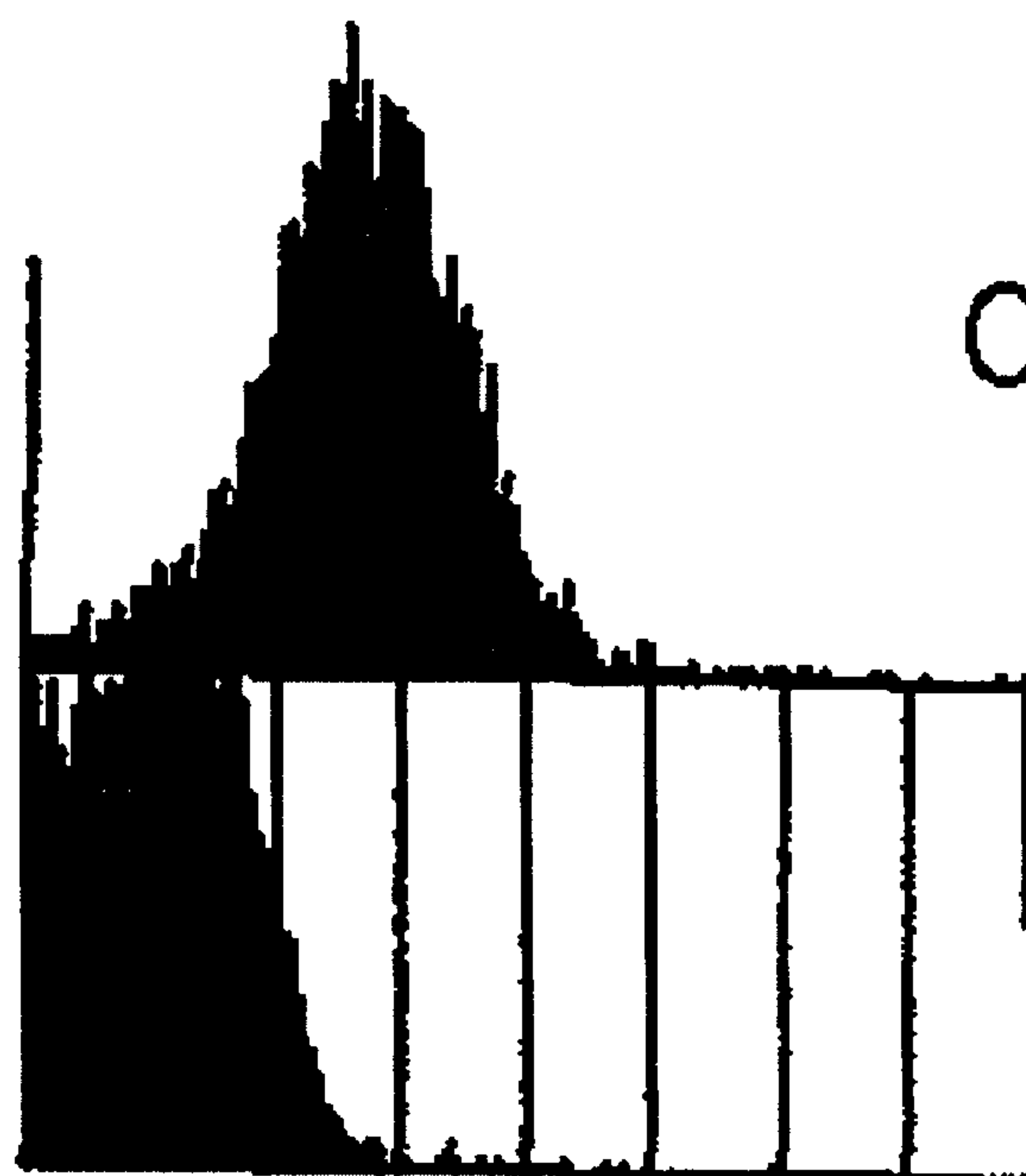
24/24

FIGURE 20**Fig. 20A**

M1 ORF132

**Fig. 20B**

M2 ORF132

**Fig. 20C**

ORF132