

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
15 April 2004 (15.04.2004)

PCT

(10) International Publication Number
WO 2004/031236 A2

(51) International Patent Classification⁷: **C07K 14/705**

(21) International Application Number:
PCT/GB2003/004273

(22) International Filing Date: 1 October 2003 (01.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0222764.3 2 October 2002 (02.10.2002) GB

(71) Applicant (*for all designated States except US*): **UNIVERSITY OF BRISTOL** [GB/GB]; Senate House, Tyndall Avenue, Bristol BS8 1TH (GB).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **VIRJI, Mumtaz** [GB/GB]; Department of Pathology and Microbiology, School Of Medical Sciences, Univesity Of Bristol, University Walk, Bristol BS8 1TD (GB).

(74) Agent: **MCQUEEN, Andrew, Peter**; Stevens Hewlett & Perkins, 1 St. Augustine's Place, Bristol BS1 4UD (GB).

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

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

Published:

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

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

(54) Title: THERAPEUTIC PEPTIDES

(57) **Abstract:** The present invention provides a ligand isolated from *Moraxella catarrhalis* outer membrane protein which binds to CEACAM receptors, said ligand comprising a receptor binding domain comprising an amino acid sequence selected from the group disclosed, or a fragment, homologue, functional equivalent, derivative, degenerate or hydroxylation, sulphonation or glycosylation product or other secondary processing product thereof. The invention also provides medicaments and vaccines comprising said ligand, and their use in the treatment or prophylaxis of infection. Also provided is a screening method for the identification of novel therapeutic compounds.



WO 2004/031236 A2

THERAPEUTIC PEPTIDES

This invention relates to therapeutic peptides and in particular to therapeutic peptides which are useful in the preparation of vaccines or other treatments for infection as well as in the screening of compounds for potential pharmaceutical activity in the treatment of infections.

The present inventors have previously identified that Carcinoembryonic antigen related cell adhesion molecules (CEACAMs) are receptors for pathogens of mucosal membranes, especially for respiratory pathogens such as *Neisseria meningitidis*, *Haemophilus influenzae* and *Moraxella catarrhalis*. CEACAMs belong to the CarcinoEmbryonic Antigen (CEA) family, a member of the Immunoglobulin superfamily. The CEA gene family comprises surface expressed (CEA) and secreted (pregnancy-specific glycoprotein, PSG) sub-families. The membrane-associated sub-family redefined as CEACAM (CEA-related cell adhesion molecule)²⁰ comprises several related glycoproteins of which CEACAM1 is the most widely expressed in distinct human tissues¹². The studies reported by the inventors primarily used Chinese Hamster Ovary (CHO) cells transfected with CEACAM1 (previously termed CD66a and BGPc) that contains four extracellular domains, a TM region and a short (S) or a long (L) cytoplasmic tail (molecular formula: NA1BA2-TM-S or L). In addition, soluble truncated constructs containing one or more of the extracellular domains were used. Previous studies demonstrated that both *Neisseria meningitidis* and *Haemophilus influenzae* primarily target the N-domain of several CEACAMs^{7,9,10}. Such targeting may lead to cell surface attachment as well as cellular invasion⁹. In addition, bacteria may bind to CEACAMs on phagocytic cells and T and B lymphocytes. Such interactions may lead to bacterial cell death⁸, target cell death or inhibition of immune function, e.g. of T and B lymphocytes when *N. gonorrhoeae* (closely related to *N. meningitidis*) binds to CEACAMs of these lymphocytes^{21,22}.

The presence of CEACAM-binding ligands in *Moraxella catarrhalis* and *Haemophilus influenzae* has been a surprising find to the present inventors since CEACAMs have long been associated with outer membrane opacity-

associated Opa proteins of *Neisseriae* and neither *Haemophilus influenzae* nor *Moraxella catarrhalis* produce Opa proteins. In the description which follows, the present invention will be described with particular reference to infection of the mucosal membranes, especially of the respiratory membranes, or to infection of the ear (especially otitis media) but it is to be understood that the invention finds equal utility in other areas such as the genital mucosa or the urethrae where CEACAM receptors are implicated in infection or other receptor-binding processes or elsewhere in human infection where bacteria may become disseminated from mucosal surfaces.

The mucosal pathogens *Neisseria meningitidis* (Nm), *Haemophilus influenzae* (Hi) and *Moraxella catarrhalis* (Mx) are human specific organisms and reside in the upper respiratory tract from where they may disseminate to cause serious infections. Meningococcal strains of distinct serogroups may be carried within the nasopharynx of up to 25% of healthy individuals¹. However, in a number of subjects, the organism invades the mucosal barrier to cause one of the most rapidly advancing and extremely serious diseases. The precise factors that increase host susceptibility to meningococcal infection are not fully understood. Moreover, the limited protection afforded by group-specific vaccines and the non-immunogenicity of the group B polysaccharide underscore the need for fundamental studies to understand host susceptibility and identify salient sub-capsular features that could serve as common targets to combat meningococci. Studies by the present inventors have provided an understanding of the molecular basis of meningococcal colonisation, the nature of its interactions with human barrier cells (epithelial and endothelial) as well as phagocytic cells. In recent years, the basis for mucosal colonisation by commensal *Neisseriae* has been investigated to understand the features which differentiate between largely harmless colonisers and occasional but serious pathogens such as Nm. In addition, the studies have determined whether commensal *Neisseriae* could be used as carriers of potential vaccine antigens of Nm.

Up to 75% of healthy individuals may carry strains belonging to the species *H. influenzae* ². Although as a result of the Hib vaccine, there has

been a dramatic decrease in the incidence of the type b disease in the West, diseases caused by non-typable Hi (NTHi) strains remain a major problem. NTHi cause localised as well as disseminated infections including epiglottitis, otitis media, cellulitis, pneumonia, endocarditis, bacteraemia and meningitis. Otitis media is one of the major problems in paediatric medicine and NTHi are responsible for over 20% of episodes in children during the first year of life ^{2,3}. NTHi are also associated with acute recurrent and persistent infections in patients with chronic obstructive pulmonary disease (COPD) and cystic fibrosis. What determines recurrent infections by NTHi in these patients or multiple episodes of otitis media in children is unclear ^{2,3,4}.

Moraxella catarrhalis, another resident of the human respiratory tract, is often isolated from cases of localised infections together with Hi. Both organisms are associated with sinusitis and exacerbations of asthmatic conditions ^{5,6}. Mx is the third most common cause of otitis media in children (estimated to be responsible for 3-4 million cases annually). It also causes lower respiratory tract infections in adults especially in patients with COPD ⁵. On rare occasions, it has been associated with disseminated infections ⁵. Both Hi and Mx cause persistent infections and are believed to escape host immune mechanisms and antibiotics by tissue penetration ⁴. Several outer membrane proteins of Mx have been studied with respect to their adhesive properties. However, few cellular receptors for Mx have been identified and many of the details of pathogenic mechanisms remain to be investigated ^{4,5,6}.

A primary requirement for respiratory mucosal pathogens is establishment of firm contact with respiratory epithelial cells. The targets of these human tropic pathogens are human specific molecules and studies have to rely on *in vitro* human tissue and organ cultures. The attachment is often mediated by bacterial phase- and antigenically-variable structures. In addition, it is becoming increasingly clear that attachment of pathogens is multi-faceted and environmental adaptation plays a significant role in the manner of attachment. Although many recent studies have begun to define various stages in the complex cellular targeting mechanisms, the details of environmental adaptation or host-microbial cross-talk remain to be described.

Recent studies by the present inventors have shown that Nm⁷⁻⁹ and Hi^{10,11} share some distinct and other common mechanisms of targeting certain human cell surface receptors CEACAMs¹².

Furthermore, the present inventors have recently identified that clinical isolates of *Moraxella catarrhalis* also target the human CEACAM molecules. In addition, it has now been found that a Moraxella outer membrane protein of high molecular weight binds to the receptor. Expression of CEACAMs in distinct tissues¹² including respiratory epithelial cells has been demonstrated²⁴. These observations imply that specific targeting of CEACAMs is of particular advantage to respiratory bacteria and may have arisen as a result of convergent evolution.

Amongst the adhesion factors elaborated by *Neisseria meningitidis* are pili (fimbriae)^{13,14,16} and the outer membrane opacity proteins, Opa and Opc^{15,16}. Nm pili are long filamentous protein structures composed of multiple pilin subunits. They are generally regarded as the most important adhesins in capsulate bacteria^{13,14,16} due the fact that capsule partly or totally masks outer membrane ligands resulting in their reduced functional efficacy, whilst pili traverse the capsule and remain functional in fully capsulate bacteria. Opa are antigenically variable family of proteins and occur in *N. meningitidis* as well as *N. gonorrhoeae*. In meningococci, 3 - 4 opa gene loci code for related transmembrane proteins with four surface exposed loops, three of which undergo sequence variation^{16,17}. Opc, another trans-membrane protein, is largely invariant^{15,16}. Over the last 12 years, the present inventors have investigated structure/function relationships of Nm pili, the virulence potential of Opa and Opc proteins and identified two human receptors for the neisserial opacity proteins. Further, the role of surface sialic acids in bacterial interactions with human target cells as well as the role of LPS and other factors in cellular toxicity have been studied by the present inventors.

The present invention results from the identification by the inventors of a ligand of high molecular weight isolated from Moraxella outer membrane protein which binds to CEACAM receptors.

The ligand can be characterised by its SDS-PAGE migration pattern which is indicative of the USP-family of proteins in that it is broken down to monomers having a molecular weight of between approximately 60 and 150 kD when boiled for a prolonged period.

The ligand has been characterised in Mx strain ATCC 25238 (MX2) as UspA1 and its amino acid sequence determined. The ligand has been further characterised to determine the receptor binding region or domain, *i.e.* peptide or peptide-associated features that bind to the receptor.

Accordingly, the present invention provides a ligand isolated from *Moraxella catarrhalis* outer membrane protein which binds to CEACAM receptors, said ligand comprising a receptor binding domain comprising an amino acid sequence selected from the group consisting of residues 463 to 863, 527 to 668, 527 to 863, 427 to 623, 427 to 668, and 427 to 863 of the sequence shown in Figure 6, or a fragment, homologue, functional equivalent, derivative, degenerate or hydroxylation, sulphonation or glycosylation product or other secondary processing product thereof.

The term ligand is used herein to denote both the whole molecule which binds to the receptor and any part thereof which includes the receptor binding domain such that it retains the receptor binding property. Thus "ligand" encompasses molecules which consist only of the receptor binding domain *i.e.* the peptide region or regions required for receptor binding.

In one embodiment, the present invention provides a ligand consisting of an amino acid sequence selected from the group consisting of residues 463 to 863, 527 to 668, 527 to 863, 427 to 623, 427 to 668, and 427 to 863 of the sequence shown in Figure 6, or a fragment, homologue, functional equivalent, derivative, degenerate or hydroxylation, sulphonation or glycosylation product or other secondary processing product thereof.

Structurally and/or functionally equivalent receptor binding domains may also occur in other UspA-like proteins since hybrid proteins occur in Mx that may contain mosaic epitopes derived from both UspA1 and UspA2 proteins²³. Ligands comprising such equivalent receptor binding domains are also within the scope of the invention.

Preferably the ligand or receptor binding domain is suitable for use in the prevention or treatment of infection.

The present invention also provides a nucleic acid sequence encoding the ligand protein of the present invention together with homologues, polymorphisms, degenerates and splice variants thereof.

The ligand of the invention, or combinations thereof may be used in a vaccine or other prophylactic treatment of infection.

The vaccine or other prophylactic treatment may comprise any known adjuvant, vehicle, excipient, binder, carrier, preservatives and the like, to provide a pharmaceutically acceptable preparation of the ligand for use in the treatment of a patient.

The invention also provides a pharmaceutically acceptable preparation of the ligand for use in medicine.

The pharmaceutically acceptable preparation of the ligand may be used in the treatment or prevention of any disease where CEACAM receptors are implicated, for example in the treatment or prevention of infection, respiratory disease, neoplastic diseases and associated conditions of neoplastic diseases, and angiogenesis.

Preferably, where an infection is treated, the infection is of, or has occurred via, the mucosal membrane, especially a respiratory infection.

Most preferably, the ligand is used as a vaccine for or in other prophylaxis or treatment of *Neisseria meningitidis*, *Haemophilus influenzae* and *Moraxella catarrhalis*. Ideally, the ligand is used as a vaccine for or in other prophylaxis or treatment of otitis media.

In a further aspect, the ligand of the present invention may also be used to identify novel blocking reagents for use as therapeutic agents to protect vulnerable groups and the public in general against several mucosal pathogens. For example the ligand may be used to identify receptor analogs which are useful for this purpose.

Hence the present invention also provides a screening assay for the identification of novel blocking reagents for use as therapeutic agents, the assay comprising the steps of screening potential therapeutic agents for their

ability to mimic or for their homology to the ligand of the present invention. The invention further provides therapeutic agents identified by the aforementioned screening assay.

Effective vaccine components may be produced by using the information of the receptor targeting mechanisms identified by the present invention such as biologically active peptide mimics. These could prevent bacterial colonisation / invasion of mucosa as well as elicit antibodies which may be blocking, opsonic and bactericidal.

Bacterially derived biologically active peptide sequences identified by the ligand of the present invention could be used to study the roles of CEACAMs in cancer and development as the molecules are associated with these processes. These also have potential as anti-cancer agents and to control or otherwise in the treatment of angiogenesis.

In a further embodiment, the invention provides the use of a CEACAM receptor-binding ligand in the manufacture of a medicament for the treatment or prophylaxis of a disease in which CEACAM receptors are involved in cellular targeting of the pathogen which causes the disease, wherein the ligand comprises an amino acid sequence selected from the group consisting of residues 463 to 863, 527 to 668, 527 to 863, 427 to 623, 427 to 668, and 427 to 863 of the sequence shown in Figure 6, or a fragment, homologue, functional equivalent, derivative, degenerate or hydroxylation, sulphonation or glycosylation product or other secondary processing product thereof.

Preferably, the disease is selected from the group consisting of infection, respiratory disease, neoplastic disease and associated conditions of neoplastic disease, and angiogenesis.

Medicaments as described above are of particular utility where the pathogen infects, or enters via, a mucosal membrane.

Medicaments as described herein are particularly useful in the treatment or prevention of infections (disease) caused by *Moraxella catarrhalis*. However, ligands as described herein are useful in the manufacture of medicaments for the treatment of any disease where CEACAM receptors are

implicated such as diseases caused by *Neisseria meningitidis* and *Haemophilus influenzae*.

In a particularly preferred embodiment, the disease is otitis media.

Ligands of the invention may also be used in the treatment of diseases caused by other oral bacteria, such as dental caries.

Embodiments of the present invention will now be described purely by way of non-limiting example in which reference is made to the figures of which:-

Figure 1 is a graph showing relative binding levels of CEACAM1-Fc ($1\mu\text{g.ml}^{-1}$) soluble receptor construct alone (white, left hand bars), or in the presence of the CEACAM1 N-domain specific antibody YTH71.3 (grey, centre bars) to 3 strains of Mx immobilised on nitrocellulose. CD33-Fc binding in each case was negligible (black, right hand bars). Strains 1, 2, 3: MX2 (ATCC 25238), MX3, MX4 (clinical isolates) respectively. Binding was determined in a dot blot overlay and the intensity of reactions quantified by densitometric analysis using NIH Scion Image program.

Figure 2 shows a Western blot of the strain MX2 proteins separated under undissociating (unheated, lane 1) or after boiling for 10 min. (lane 2). Blot was overlaid with CEACAM1-Fc ($1\mu\text{g.ml}^{-1}$) and the receptor binding detected with anti-human Fc antibody conjugated to horseradish peroxidase and its substrate.

Figure 3 shows Western blots of denatured whole cell lysates of strains MX2, -3, -4 (lanes 1-3 respectively) were overlaid with CEACAM1-Fc ($1\mu\text{g.ml}^{-1}$; a), anti-UspA1 peptide antibody ($10\mu\text{g.ml}^{-1}$; b) and anti-UspA2 peptide antibody ($10\mu\text{g.ml}^{-1}$; c). Note the similar migration profile of CEACAM1-Fc binding proteins and anti-UspA1 binding proteins in the three strains. Remnants of undissociated proteins at c. 250 kDa, (detected in this case due to higher sensitivity of detection in the alkaline phosphatase assay), bind to CEACAM1-Fc. These are only weakly recognized by the anti-peptide antibodies, presumably since the epitopes contained within synthetic peptides are not fully exposed in the native protein, which become progressively

exposed as the complex denatures. Anti-UspA2 antibodies bind to proteins of apparent masses > 200 kDa, which remain undissociated after boiling, a property indicative of UspA2 proteins.

Figure 4 shows a mass spectrum of tryptic peptides of the CEACAM1 binding protein of MX2 following electro-elution (4a). The summary of data input into the ProFound protein identification database is shown in (4b). A table of the top ten identified proteins and the probability values are shown in (4c). Detail of the number 1 ranked candidate, in this case UspA1 with a Z-score of 2.34 is shown in (4d) indicating the number of peptides matched, their positions within the protein, the % of the protein covered and a list of peptides unmatched on this occasion.

Figure 5 shows Western blot analysis of tryptic fragments of UspA1 of MX2. A: control blot using secondary antibodies (mixture of goat anti-human Fc and goat anti-rabbit Ig used in B and C). B: blot overlaid with CEACAM1-Fc and goat anti-human Fc. C: blot overlaid with affinity purified rabbit antibodies raised against UspA1 peptide (ETNNHQDQKIDQLGYALKEQGQHFNRR) (see Figure 6) and anti-rabbit Ig. * = lanes with molecular weight markers – shown on the left. The peptides shown by the double arrows react strongly with CEACAM1-Fc (B) as well as the anti-UspA1 peptide antibodies (C). The binding of the anti-UspA1 peptide antibodies to the lowest MW peptide (arrowhead) identifies this CEACAM-binding fragment as the C-terminal fragment contained within N-199 to K-863 of UspA1 of MX2 (see Figure 6).

Figure 6 shows the amino acid sequence of MX2 UspA1 protein. UspA1-specific peptide used to raise antisera in rabbits is shown in bold. The CEACAM-binding region is contained in the underlined C-terminal fragment of UspA1 of MX2.

Figure 7 shows the separation of tryptic peptides reacting with CEACAMs. *M. catarrhalis* strain MX2 was treated with 1mg/ml trypsin at 37°C for 10 min. Trypsinised sample was subjected to SDS-PAGE. After staining, 50 kDa region was electroeluted overnight. Electroeluted protein was freeze dried and resuspended in buffer and applied to a second gel. Part of the gel was

blotted onto nitrocellulose and peptide bands reacting with CEACAM were identified by Western blot overlay using CEACAM1-Fc (Blot).

‘*’: denotes peptide bands sent for N-terminal sequencing.

Figure 8 shows the amino acid sequence of a tryptic peptide of MX2 UspA1 protein. The 50 kDa tryptic peptide shown (amino acids 463 to 863) binds CEACAM and antiserum against UspA1 peptide (amino acids 753-780 underlined). The N-terminal sequence of the c. 50 kDa CEACAM binding peptide is “ALESNVEEGL” that occurs after the trypsin cleavage site at amino acid 462.

Figure 9 shows a diagrammatic representation of the positions of primers used to generate *uspA1* gene fragments for the expression of recombinant peptides. The CEACAM1 binding site was encoded by DNA amplified by primers P4 and P7, additional primers throughout this region (letters A-I) were designed and employed.

Figure 10 shows the sequence of the recombinant fragment 4-7. The underlined region is the N-terminal region of the CEACAM1-reactive tryptic peptide. The predicted molecular weight of the fragment 4-7 is c. 26 kDa. The predicted MW of the His-tagged fragment: c. 28kDa. The position of the truncated peptide is shown by ‘T’ (see Figure 13).

Figure 11 is a diagram showing the general cloning, expression and purification strategy for recombinant UspA1 peptides, as exemplified by the pQE30 system.

Figure 12 is a map of the vector pQE30.

Figure 13 shows binding of CEACAM1-Fc in blot overlay to recombinants 4-8 (lane 3), 4-T (lane 4) and 4-7 (lane 5) polypeptides. Lane 1 contained a Treponemal control recombinant peptide and lane 2 contained lysates of non-induced M15 containing 4-8 construct.

Figure 14 shows Western blots showing recombinant peptide reactivity with anti-His tag antibody (top) and CEACAM1-Fc (bottom). Lanes 2-4 contained 6-8 peptide, lane 1 contained 4-8 as a control. Predicted migration positions of the peptides are shown on the right. Both peptides bind anti-His antibody. However, whilst 4-8 binds to CEACAM1-Fc, 6-8 does not.

Figure 15 shows that D-8 polypeptide binds CEACAM1-Fc (lane 1) but not CD33-Fc used as a control (lane 2). The origin of this peptide was confirmed by reactivity with the anti-His tag antibody (lane 3).

Figure 16 is a schematic diagram showing the relative sizes and positions of rUspA1 fragments. Recombinant 4-7 has been used for bacterial blocking – see Figure 17.

Figure 17 shows CHO-CEACAM1 transfectants were incubated in the absence of peptides (A) or with recombinant control peptide (a treponemal peptide, B) or UspA1 r4-7 peptide (sequence corresponding to the strain MX2, C and D) at the concentrations shown and bacteria added for a period of 2 hours. At the end of this incubation, unbound bacteria were washed off and bound bacteria detected with anti- *M. catarrhalis* polyclonal antiserum and TRITC conjugated secondary antibody. At 1 µg/ml significant inhibition of a heterologous *M. catarrhalis* strain (MX1) was obtained and at 10 µg per ml, *H. influenzae* binding was significantly inhibited by *M. catarrhalis* UspA1 recombinant peptide.

Example 1. *Moraxella catarrhalis* strains bind to human CEACAM1-Fc via the UspA1 proteins

The strains of *Moraxella catarrhalis* (Mx) used in this study include clinical isolates (MX3 and MX4) as well as reference strain purchased from American Type Culture Collection (ATCC 25238, a clonal culture of this was designated MX2).

Receptor overlay assays show that Mx bind to CEACAM1-Fc receptor constructs

To assess the interaction of *M. catarrhalis* strains with CEACAM1, bacteria (c. $4 - 8 \times 10^6$) were applied to nitrocellulose, air dried and non-specific binding sites blocked in 3 % BSA-PBST. Nitrocellulose strips were overlaid with either CEACAM1-Fc ($1-2 \mu\text{g ml}^{-1}$) alone or in the presence of the CEACAM1 N-domain antibody YTH71.3. CD33-Fc ($1-2 \mu\text{g ml}^{-1}$) was used as a negative control. Chimeric protein constructs were prepared as previously

described¹¹. Chimeric receptor binding was detected by goat anti-human Fc antibodies conjugated to either horseradish peroxidase or alkaline phosphatase. Blots were developed with either diaminobenzidine and hydrogen peroxide or nitroblue tetrazolium and 5-bromo-4-chloro-3-indoylphosphate substrates respectively. All Mx strains bound to CEACAM1-Fc but not to CD33-Fc (Fig.1). The receptor binding was inhibited in the presence of the monoclonal antibody YTH71.3 suggesting that the strains bound to the N terminal domain of the receptor. Accordingly, all the strains bound equally well to the truncated N-Fc construct of CEACAM1, containing only the N domain (not shown).

Identification of CEACAM1-Fc binding protein of *M. catarrhalis*

Whole cell lysates of Mx (c. 3×10^7) were applied to individual lanes of a 10% bis-tris polyacrylamide gel (Invitrogen) either without previous heat treatment or after heating at 100°C for 10 minutes and subjected to electrophoresis for 45 min at 180 V. Proteins from the gels were transferred to nitrocellulose membranes using standard blotting conditions. Nitrocellulose membranes were overlaid with soluble chimeric constructs as described above. In each case, a single protein was observed that bound specifically to CEACAM1-Fc. The migration of the protein on gels was indicative of UspA1 protein of Mx, since when bacterial lysates were applied without heat denaturation, the CEACAM1 binding proteins migrated with an apparent mass of > 200 kDa whereas when the lysates were first heated, the CEACAM1-binding proteins migrated with a reduced mass of *circa* 92 kDa (Fig. 2)

Antibodies raised against UspA1 but not those against a similar protein UspA2 bind to the CEACAM1-Fc binding proteins of Mx strains

Peptides were designed according to published sequences for the UspA proteins of *M. catarrhalis* strains. Namely, ETNNRQDQKIDQLGYALKEQGQHFNNR (UspA1-peptide) and KDEHDKLITANKTAIDANKAS (UspA2-peptide). Peptides were coupled to KLH via an N-terminal cysteine residue that was incorporated and were used

to immunise rabbits (200 µg peptide per rabbit) at 14 day intervals, initially with complete Freund's adjuvant and thereafter with incomplete Freund's adjuvant. Rabbits were bled at day 0 and at 14 day-intervals post immunisation. Polyclonal antibodies were purified using the appropriate peptide coupled to AminoLink Plus columns (Pierce). UspA-specific antibodies were used at a concentration of 1-10 µg ml⁻¹ for Western blot overlays and detected with goat anti-rabbit secondary antibody coupled to alkaline phosphatase. Blots were developed as described earlier. The migration of the CEACAM1-Fc binding proteins on SDS-PAGE of the three Mx strains was identical to that of the UspA1-antibody binding proteins but not UspA2-antibody binding proteins (Fig. 3)

M. catarrhalis ligand co-precipitating with CEACAM1-Fc is identified as UspA1

Overnight cultures of bacteria were suspended in 100 mM octyl βD glucopyranoside in PBSB containing a protease inhibitor cocktail (pic; PMSF 1 mM, E-64 1 µM, pepstatin A 1 µM, bestatin 6 nM and EDTA 100 µM). Samples were mixed end over end overnight at 4°C. Meanwhile 100 µl protein A coupled to sepharose CL-4B (Sigma) was incubated with either, 20 µg CEACAM1-Fc or CD33-Fc (used as a control) overnight at 4°C and subsequently washed 3 times with PBSB to remove any unbound receptor. Insoluble bacterial material was removed by centrifugation at 15,000 g for 30 min. Soluble extract was incubated with either receptor-Protein A sepharose complex for 2h at 4°C (at a ratio of 5 x10⁸ bacteria per µg of receptor construct). Following extensive washing with 50 mM octyl βD glucopyranoside and PBSB samples were analysed by SDS-PAGE electrophoresis and Western Blotting under denaturing conditions.

In co-precipitation experiments with CEACAM1-Fc, MX4 yielded a strongly staining protein of c. 97 kDa and MX3 yielded a relatively weakly staining protein of c. 92 kDa. The masses of co-precipitated proteins corresponded to those observed in the receptor overlay experiments (shown in Fig. 3). Neither protein co-precipitated with CD33-Fc. The co-precipitated

proteins were further identified as UspA1 proteins since they bound to anti-UspA1 peptide antibody. In addition, following excision from the gel, the MX4 protein was subjected to MALDI-TOF mass spectrometry (see below).

CEACAM1 ligand identification by MALDI-TOF mass spectrometry

(a) Western overlay samples. Whole cell lysates of MX2 and MX3 were subjected to SDS-PAGE in trench gels and the protein band corresponding to CEACAM1-Fc binding ligand was electroeluted from the gel. The sample was concentrated and reapplied in a single lane of a second gel, subjected to electrophoresis prior to in-gel trypsin digestion of the appropriate protein. The resulting peptides were analysed by Matrix-assisted Laser desorption/ionisation–time-of–flight (MALDI-TOF) mass spectrometry. An example of the resulting mass spectrum is shown for the CEACAM binding protein of MX2 (Fig 4a). The peptide masses obtained were entered into the ProFound protein identification site and the results were obtained as shown for this protein (Fig. 4b, c). In this case 10 peptide masses were matched to predicted masses for tryptic peptides of UspA1 of *M. catarrhalis* covering approximately 18 % of the protein (Fig. 4d). The estimated Z score of 2.34 is strongly suggestive of the protein being UspA1 (Z score >1.65 are above the 95th percentile; http://129.85.19.192/profound_bin/webProFound.exe). In addition, another non-binding high molecular weight band of MX2 was identified as UspA2 following a similar analysis. Similarly for strain MX3, the CEACAM1 binding and non-binding proteins were identified as UspA1 and UspA2 respectively.

(b) Co-precipitated samples: For MX4, the CEACAM1-Fc co-precipitated protein (as described above) was also identified as UspA1 by MALDI-TOF MS with a Z score of 2.27 in an all taxa search. 12 peptides were matched covering 21% of the protein.

This study therefore has identified that *Moraxella catarrhalis* targets human CEACAM1 via the high molecular weight protein UspA1 in the reference and clinical strains indicated.

Enzymatic cleavage of UspA1 and recombinant peptide

(a) Tryptic peptides of UspA1 bind to CEACAM1-Fc

MX2 bacterial suspensions (10^{10} ml⁻¹) were treated with Trypsin (Sigma) at 0.1-1 mg/ml concentrations and incubated for 1-4 hours at 37°C. Digested lysates were dissociated in SDS-PAGE buffer, boiled and subjected to electrophoresis. After transfer to nitrocellulose, the blots were overlaid with the receptor construct CEACAM1-Fc or the affinity purified anti-UspA1 peptide antibodies. A small fragment reacting with the receptor also bound to the anti-UspA1-specific antibodies (Fig. 5).

(b) Localisation of CEACAM binding domain of UspA2 of M. catarrhalis strain MX2

MX2 bacterial suspensions were treated with Trypsin at 1mg/ml for 10 minutes at 37°C.

The trypsinised sample was run on an SDS-PAGE gel. After staining, the 50 kDa region was electroeluted overnight. Electroeluted protein was freeze dried and resuspended in buffer and applied to a second gel. Part of the gel was blotted onto nitrocellulose and peptide bands reacting with CEACAM were identified by Western blotting overlay using CEACAM1-Fc. (Fig.7).

Peptide bands corresponding to circa. 50 kDa and circa. 150 kDa peptides were subjected to N-terminal sequencing. The N-terminal sequences were ALESNVEEGL (c. 50 kDa peptide) and ALESNV (c. 150 kDa peptide). The 150 kDa protein is apparently a trimer of the 50 kDa protein as they have the same N-terminal sequence. The N-terminal sequence of this peptide of MX2 UspA1 is shown in Figure 8.

(c) Recombinant peptide

The N-terminal recombinant MX2 peptide that was constructed consisting of amino acids 1 to 449 of the sequence shown in Figure 6 does not bind CEACAM. This further indicates that the c. 50 kDa tryptic peptide consisting

of amino acids 463 to 863 of the sequence shown in Figure 8, having the N-terminal sequence ALESNVEEGL contains the CEACAM-binding domain.

Example 2. Identification of Receptor Binding Domains on Multiple Virulence Determinants of Mucosal Pathogens

N. meningitidis and *H. influenzae* target human CEACAM molecules via ligands that bind to overlapping sites on CEACAM. The Mx ligand/s of the present invention also target the N domain. These observations point to an exciting possibility that similar features on ligands of several mucosal pathogens may be involved in receptor targeting.

Since Nm and Hi interactions with CEACAMs are affected by the structural features of the surface expressed variable domains, this suggests that they may contain similar crucial amino acids in a spatial configuration that are able to bind to the receptor and these may be conserved in otherwise variable proteins. Indeed our studies and those of others suggest that the two hyper-variable loops of Opa (HV1 and HV2) may be involved in CEACAM targeting^{9,18}. One powerful technique that can identify possible determinants of interactions between ligands and receptors is phage display, which can be used to identify mimotopes (random sequences to mimic the binding domains)¹⁹ as well as aptamers (sequences more closely related to the original structure that inhibit ligand binding)²⁰.

The current invention also makes it highly feasible that the Mx ligand domain that binds CEACAMs will act as a mimic for other CEACAM-binding mucosal pathogens and the structural features of this ligand will help identify the salient features required in Nm and Hi ligands for CEACAM N-domain targeting. Antibodies to the Mx domain could have the potential of identifying other ligands that have the capacity to target the same, similar or closely positioned region of the receptor.

Identification of the minimal CEACAM1 binding domain of MX2 UspA1 is being undertaken using known methods of protein engineering and recombinant DNA technology. Recombinant peptides can be screened in vitro by receptor overlay assays described above to detect the domain of MX2 that

binds to the receptor. His-Tagged peptides can be separated on a nickel column, His-Tag cleaved as required and used for immunising rabbits and mice to obtain antibodies for further investigations.

A peptide of suitable length for biological applications may be determined by examining immunological stimulatory properties as well other functions such as blocking of receptor binding.

Example 3. Salient Features of the Receptor Required for Ligand Interactions

Since the N-domain of CEACAM is sufficient for the interactions of Opa proteins, the present inventors are using phage display technique also to investigate adhesive epitopes of the CEACAM N-domain. The knowledge of the ligand-binding region/s on the receptor, which has been studied by the present inventors by alanine scanning mutagenesis of the receptor⁹ has facilitated this study. In this case also, a ligand overlay assay is available for biopanning (affinity concentration) of chimeric phages that bear receptor sequences.

Receptor analogues have the potential to block multiple strains independently of the Opa type produced since our studies have already shown that despite their antigenic variation, distinct Opa proteins require common features on the receptor for primary adhesion⁹. It is also of interest to note that CEA antigens are shed from the gut mucosa and may block adhesion of *E. coli* strains also known to target CEACAMs. This has been proposed as a mechanism of innate immunity vs. enteric pathogens.¹² Thus these receptor analogues act as therapeutic agents.

Example 4. Production of Recombinant CEACAM-binding *Moraxella catarrhalis* UspA1 peptides

Summary:

PCR amplification of several fragments along the length of UspA1 was carried out using the primers shown in Figure 9. Recombinant peptides were

obtained as described below. In the first round, it was found that the recombinant peptide that bound to CEACAM1 was encoded by DNA amplified by primers P4 and P8 but not that encoded by region between P1 and P5. Further, recombinant P4 - P7 bound CEACAM1 but not P6 - P8. Additional primers were used within P4 and P7 region. Region D-7 retained CEACAM-binding. The sequence of fragment 4-7 is shown in Figure 10.

General cloning, expression and purification strategy for recombinant UspA1 peptides

The UspA1 fragment 1-5 was produced using the pBAD system. The required PCR products were TA cloned into pBAD vector and TOP10 *E. coli* strain was used for amplification. The rest of the procedure was as shown in Fig. 11 with the exception that pBAD was induced using arabinose. The fragment 4-7 was produced using both the pBAD and pQE30 (Fig. 11) systems with similar CEACAM-binding results. The rest of the fragments were produced using the strategy in Fig. 11.

Vector and the pQE30 expression system

The vector pQE30 (Fig. 12) was used in conjunction with *E. coli* strain M15. M15 contains the plasmid pREP4 encoding a repressor, which restricts transcription of DNA cloned into pQE30. The addition of IPTG to a concentration of 1mM prevents coding of this repressor and thus transcription of the cloned fragments in pQE30.

The cloning strategy

PCR amplimers were first ligated into pCR2.1. This provided a stable host from which the amplimer could be recovered by restriction digest (*Bam*HI/*Pst*II), ensuring that each end of the gene fragment was cut. pQE30 was similarly digested and recovered by gel purification, which also ensured that both restriction sites had been cut. Digested pQE30 and UspA1 amplimer were ligated overnight at 16°C with T4 DNA ligase and transformed into CaCl₂ competent *E. coli* M15. Transformants were selected for on LB agar

supplemented with ampicillin (100 µg/ml) and kanamycin (25 µg/ml). 4-8 colonies were picked and grown in LB broth with antibiotics. Bacteria from 3-4 ml of culture were collected by centrifugation and subjected to the alkaline-lysis miniprep method. Purified vector was digested as above to check for *uspA1* insert. Bacteria containing pQE30 with a correct sized insert were grown in 50 ml. cultures and induced with IPTG (see below). Western blots were then used to screen for recombinant protein production. In addition, vectors were sequenced to determine that the insert was the correct region of DNA and to check for any sequence errors.

Expression and purification

M15 containing pQE30/*uspA1* constructs were grown in LB (supplemented with 100 µg/ml Ampicillin and 25 µg/ml Kanamycin) broth with shaking to OD₆₀₀ = 0.5 before the addition of IPTG to a concentration of 1 mM. Cultures were incubated for a further 3-4 hours and bacteria recovered by centrifugation. Bacterial pellets were solubilised in buffer B (8M urea, 50mM Tris, 10% ethanol, 2% Tween, 5mM imidazole, pH 7) for 1-3 hrs and membrane material removed by centrifugation at 20,000 g for 20 min. Supernatants were incubated with nickel resin for 1-2 hrs on a rotary mixer, and passed through a polypropylene column. The retained resin was washed with 5-10 ml buffer B and bound protein eluted with 0.5 ml elutions of buffer B supplemented with 100 mM imidazole. Eluted proteins were checked by SDS-PAGE before dialysis to remove urea and other salts.

Recombinant fragments

A: Fragments 1-5 and 4-8

These fragments produced by pBAD system showed that 1-5 did not bind CEACAM1 but 4-8 did.

B: Fragments 4-8, 4-8T, 4-7 and 6-8 produced using pQE30 system.

Fragment 4-8 was the first rUspA1 fragment produced and was found to bind CEACAM1 with a high affinity in blot-overlay assays. In addition to the

full-length 4-8 rUspA1 peptide, a smaller, truncated, protein was observed. This peptide (designated 4-T) also bound CEACAM1 and appeared to be expressed at lower levels than 4-8 (Fig. 13). Sequence analysis of pQE30/4-T found that a mismatch (CAA to TAA) led to a termination codon at residue Q624 (see Fig. 10).

Peptides 4-7 and 6-8 were produced in order to rule out the possibility that a second CEACAM binding site occurred in 6-8. As predicted from the 4-T peptide, 4-7 demonstrated CEACAM binding (Fig. 13) but 6-8 did not (Fig. 14).

C: Fragments D-8 and D-7 produced by pQE30 system

Both rUspA1 fragments D-8 (Fig. 15) and D-7 (not shown) were found to bind CEACAM1.

Biological activity of recombinant peptide 4-7

M. catarrhalis strain MX1 and *H. influenzae* strain Rd bind to Chinese Hamster Ovary (CHO) cells transfected with CEACAM1. Their binding can be blocked with *M. catarrhalis* UspA1 recombinant peptide 4-7 but not a control peptide (Fig. 17).

References:

1. Cartwright, K. et al. 1995. In *Meningococcal Disease*. John Wiley & Sons.
2. van Alphen, L. & van Ham, S. M. . 1994. *Rev Med Microbiol* 5:245.
3. Foxwell, A. R. et al. 1998. *Microbiol Mol Biol Rev* 62:294.
4. van Alphen, L. et al. 1995. *Am J Respir Crit Care Med* 151:2094.
5. Karalus, R. et al. 2000. *Microbes & Infection* 2:5
6. Kraft, M. 2000. *Clin Chest Med* 21:301.
7. Virji, M. et al. 1996. *Mol Microbiol* 22:929.
8. Virji, M. et al. 1996. *Mol Microbiol* 22:941.
9. Virji, M. et al. 1999. *Mol Microbiol* 34:538.
10. Virji, M. et al. 2000. *Mol Microbiol* 36:784.
11. Hill, D. et al. 2001 *Mol Microbiol* 39: 850.
12. Hammarström, S. 1999 *Cancer Biol* 9:67.
13. Stephens, D. S. 1989. *Clin Microbiol Rev* 2:S104.
14. Virji, M. et al. 1991. *Mol Microbiol* 5:1831.
15. Achtman, M. 1995 *Trends Microbiol* 3:186.
16. Merz, A. J. & So, M. 2000 *Annu Rev Cell Dev Biol* 16:423.
17. Woods, J. P. & Cannon, J. G. 1990 *Infect Immun* 58:569.
18. Wang, L-F & Yu, M. 1996 In *Methods Mol Biol* 66:269 (Morris, G. E. ed). Humana Press.
19. Colman-Lerner. 2000. *Trends Guide* p. 56 (Wilson, E. ed.).
20. Beauchemin, N., et al., 1999 *Exp Cell Res* 252: 243-249.
21. Boulton I.C. & Gray-Owen S.D. 2002 *Nat Immunol* 3(3):229-36.
22. Chen T. et al., 2001 *J Leukoc Biol* 70(2):335-40.
23. Lafontaine, E.R., et al., 2000 *J Bacteriol* **182**: 1364-1373.
24. Virji, M. 2001 *Trends in Microbiology*, 9: 258-259.

CLAIMS

1. A ligand comprising a CEACAM receptor binding domain, said ligand consisting of an amino acid sequence selected from the group consisting of residues 463 to 863, 527 to 668, 527 to 863, 427 to 623, 427 to 668, and 427 to 863 of the sequence shown in Figure 6, or a fragment, homologue, functional equivalent, derivative, degenerate or hydroxylation, sulphonation or glycosylation product or other secondary processing product thereof.
2. A medicament comprising the ligand according to claim 1 and one or more pharmaceutically acceptable adjuvants, vehicles, excipients, binders, carriers, or preservatives.
3. Use of the medicament according to claim 2 for the treatment or prophylaxis of infection.
4. Use of the medicament according to claim 2 for the treatment or prevention of any disease where CEACAM receptors are implicated.
5. Use according to claim 4 for the treatment or prevention of infection, respiratory disease, neoplastic diseases and associated conditions of neoplastic diseases, and angiogenesis.
6. Use according to any one of claims 3 to 5, where the infection is of, or has occurred via, a mucosal membrane.
7. Use according to claim 6, wherein the infection is caused by *Neisseria meningitidis*, *Haemophilus influenzae* or *Moraxella catarrhalis*.
8. Use of the medicament according to claim 2 for the prophylaxis or treatment of otitis media.

9. A screening assay for the identification of novel blocking reagents for use as therapeutic agents comprising screening potential therapeutic agents for their ability to mimic, or for their homology to, the ligand according to claim 1.
10. A vaccine comprising the ligand according to claim 1 and one or more pharmaceutically acceptable adjuvants, vehicles, excipients, binders, carriers, or preservatives.
11. A method of treating or preventing infection in an individual in need thereof comprising administering an effective amount of a medicament according to claim 2.
13. A ligand substantially as hereinbefore described with reference to and as illustrated by Figure 6 of the accompanying drawings.
14. A medicament substantially as hereinbefore described comprising the ligand of claim 13.
15. Use of a CEACAM receptor-binding ligand in the manufacture of a medicament for the treatment or prophylaxis of a disease in which CEACAM receptors are involved in cellular targeting of the pathogen which causes the disease, wherein the ligand comprises an amino acid sequence selected from the group consisting of residues 463 to 863, 527 to 668, 527 to 863, 427 to 623, 427 to 668, and 427 to 863 of the sequence shown in Figure 6, or a fragment, homologue, functional equivalent, derivative, degenerate or hydroxylation, sulphonation or glycosylation product or other secondary processing product thereof.

16. Use according to claim 15 wherein the disease is selected from the group consisting of infection, respiratory disease, neoplastic disease and associated conditions of neoplastic disease, and angiogenesis.
17. Use according to claim 15 or 16, wherein the pathogen infects, or enters via, a mucosal membrane.
18. Use according to any of claims 15 to 17, wherein the pathogen which causes the disease is selected from the group consisting of *Neisseria meningitidis*, *Haemophilus influenzae* and *Moraxella catarrhalis*.
19. Use according to claim 18, wherein the disease is otitis media.

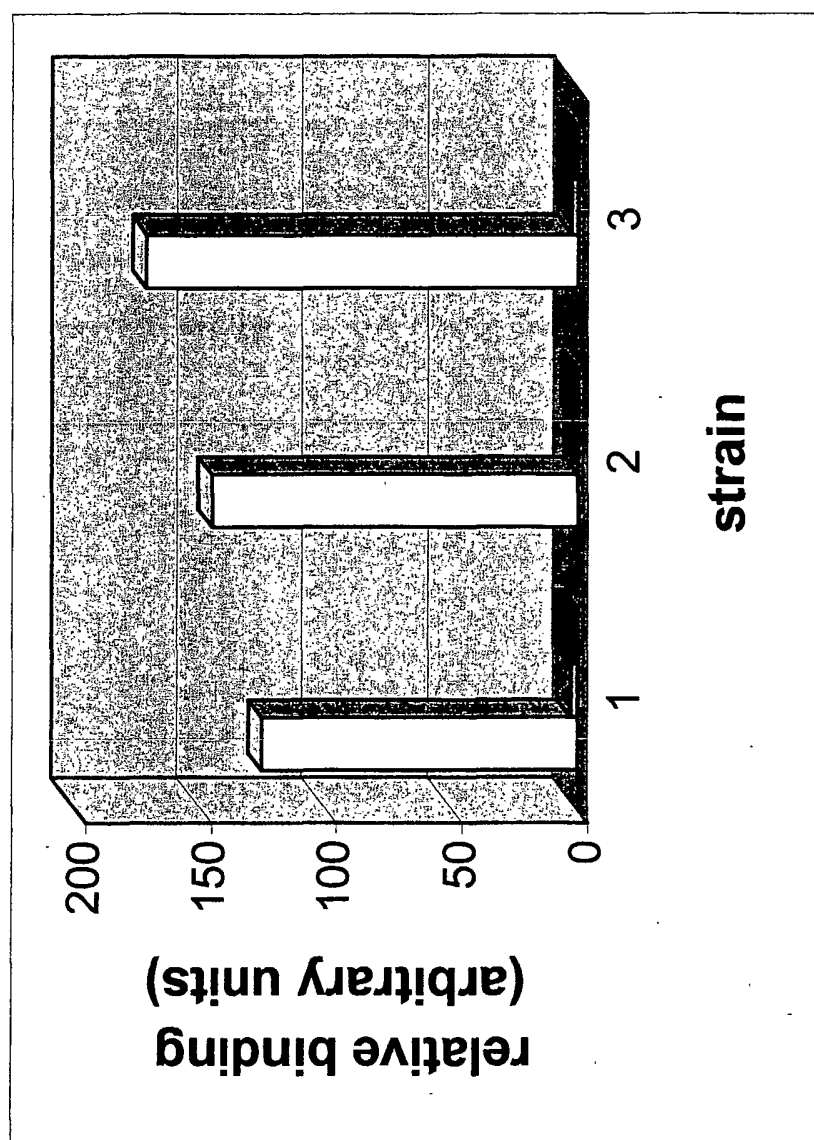


Figure 1

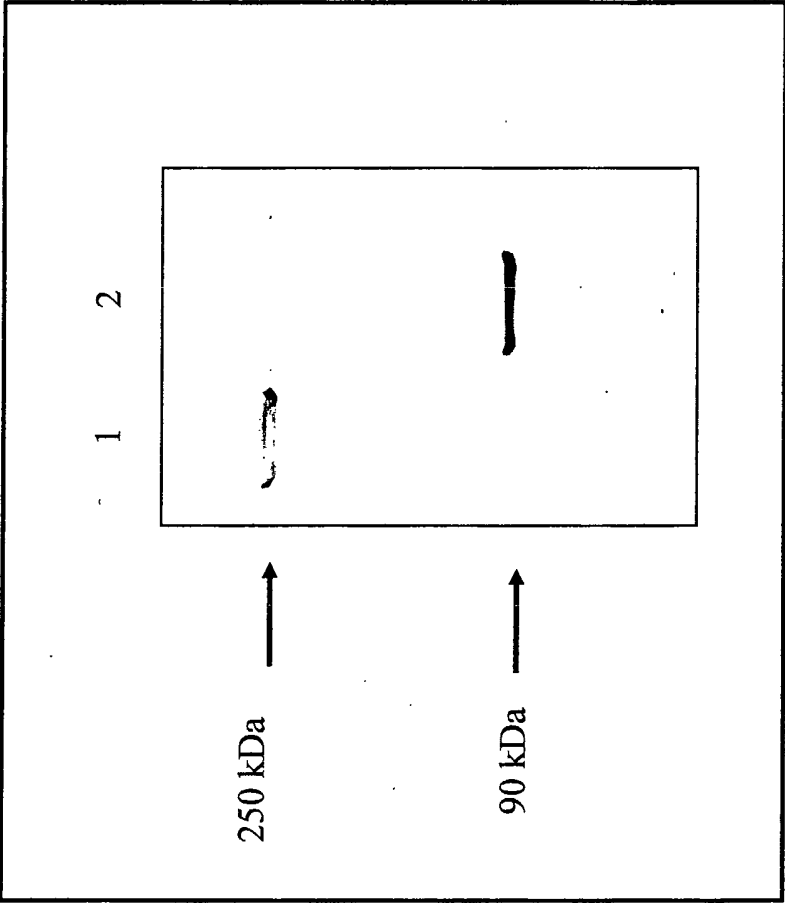


Figure 2

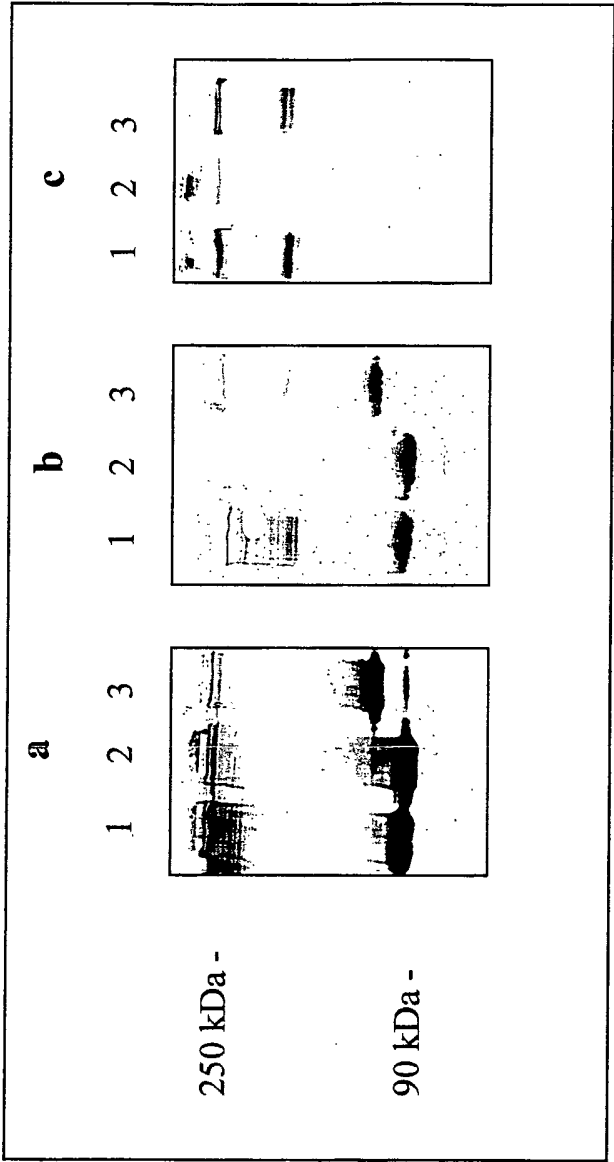


Figure 3

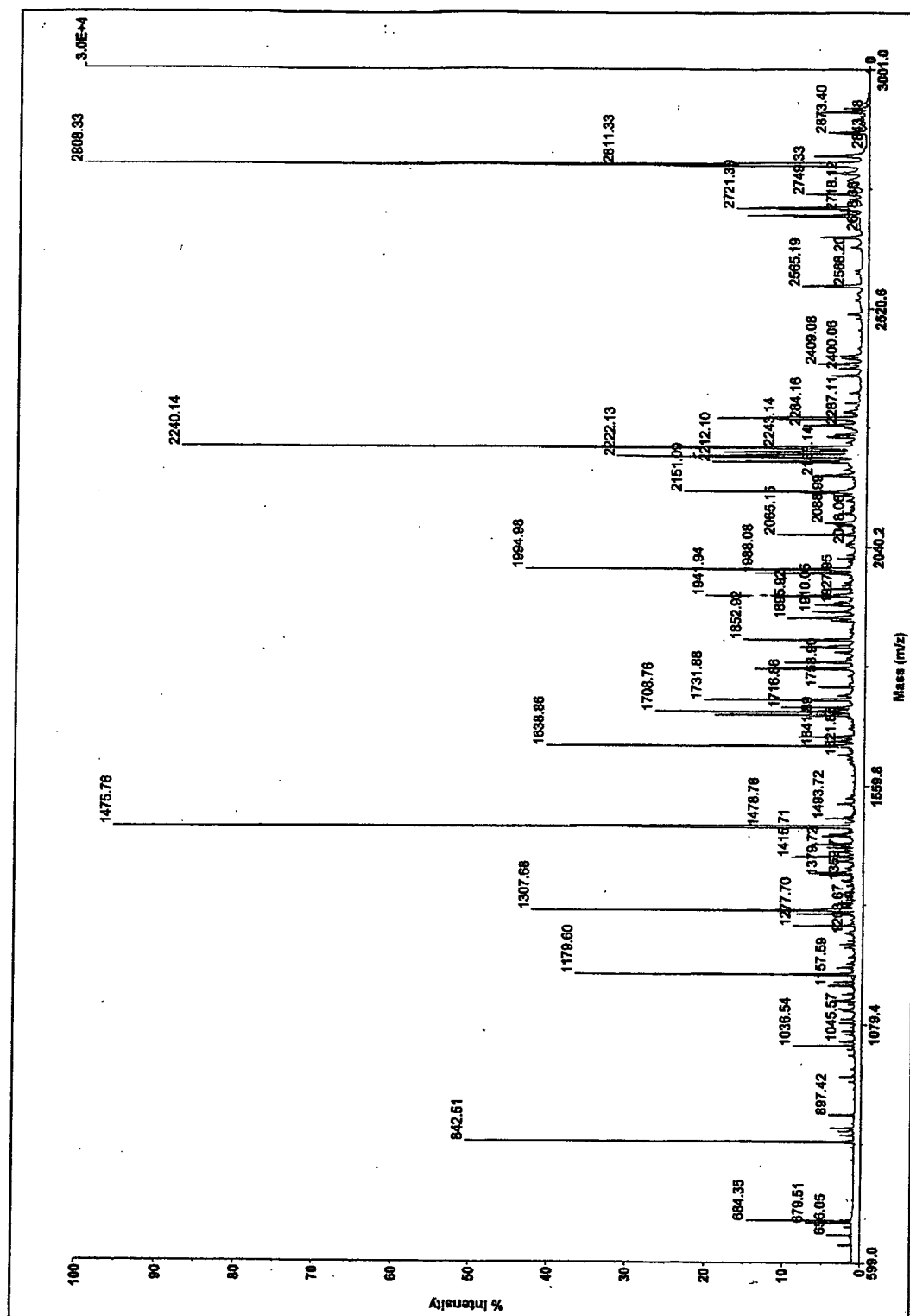


Figure 4a

Input Summary

Date & Time: Thu Aug 02 11:55:23 2001 UTC (Search Time: 17.38 sec.)

Sample ID: 11020EE20-7-01

Database: NCBItr (2001/07/29)

Taxonomy Category: All-taxa

Protein Mass Range: 80 - 100 kDa

Protein pI Range: 0.0 - 14.0

Search for: Single protein only

Digest Chemistry: Trypsin

Max Missed Cut: 1

Modifications: None

Charge State: MH+

Peptide Masses
(Da, Average)

Tolerance (AVG): 1.00 ppm

Peptide Masses: 679.514 684.348 842.510 897.421 1036.535 1126.566 1157.592 1179.598 1277.702
 (Da, Monoisotopic) 1300.538 1302.633 1306.630 1307.681 1311.689 1379.718 1383.708 1415.714
 1424.752 1434.769 1438.723 1440.774 1458.736 1474.731 1475.757 1637.850
 1638.859 1648.850 1657.807 1700.854 1701.869 1707.781 1715.823 1716.858
 1731.876 1757.875 1794.809 1807.903 1826.798 1837.968 1851.920 1890.972
 1895.922 1909.054 1922.930 1927.947 1940.931 1955.978 1987.056 1993.976
 2016.988 2064.139 2081.967 2087.974 2150.075 2184.140 2211.104 2221.129
 2225.129 2230.218 2239.136 2261.122 2283.164 2297.197 2299.182 2383.981
 2399.049 2408.087 2564.193 2663.346 2705.179 2717.129 2720.281 2748.314
 2781.924 2791.004 2792.915 2805.359 2807.327 2825.331 2872.402 2914.526
 2921.383

Tolerance (MON): 20.00 ppm

Number of Peptides: 82

Figure 4b

ProFound - Search Result Summary

The Rockefeller University Edition
Version 4.10.5

Protein Candidates for search C2BABE89:01DC-7445F3E9 [731183 sequences searched]

Rank	Probability	Est'd Z	Protein Information and Sequence Analyse Tools (T)	%	pI	kDa	R
+1	1.0e-000	2.34	T.gil5453174[gb AA043465.1 AF113606.1 (AF113606) UspA1 [Moraxella catarrhalis]	18	5.7	90.52	R
	-	-	T.gil7208237[gb AAF40122.1 AF181076.1 (AF181076) UspA1 [Moraxella catarrhalis]	14	5.4	91.86	R
	-	-	T.gil5453178[gb AAD43467.1 AF113608.1 (AF113608) UspA1 [Moraxella catarrhalis]	15	5.7	99.06	R
	-	-	T.gil5453182[gb AAD43469.1 AF113610.1 (AF113610) UspA1 [Moraxella catarrhalis]	13	5.3	95.75	R
	-	-	T.gil7208229[gb AAF40118.1 AF181072.1 (AF181072) UspA1 [Moraxella catarrhalis]	11	5.6	96.95	R
2	1.6e-011	-	T.gil8572547[gb AB96359.2 (U57551) high molecular weight outer membrane protein [Moraxella catarrhalis]	8	4.9	88.27	R
	-	-	T.gil2128720[pri E64449 hypothetical protein MJ1198 - Methanococcus jannaschii]	8	5.6	85.92	R
3	1.3e-012	-	T.gil7108403[gb AAF36416.1 U61725.1 (U61725) UspA1 [Moraxella catarrhalis]	13	5.7	93.34	R
4	9.5e-013	-	T.gil13366277[emb CAB51587.2 (AL031687) dJ998H6.1 (ortholog of rat PB-Cadherin) [Homo sapiens]	12	4.6	89.07	R
5	7.9e-013	-	T.gil11270805[pri H81749 DNA topoisomerase I IC0012 [imported] - Chlamydia muridarum (strain Nigg)	10	9.2	97.92	R
6	5.4e-013	-	T.gil9629643[ref NP_044927.1 capsid [adeno-associated virus 4]	11	6.1	80.62	R
7	4.9e-013	-	T.gil7497065[pri T15778 hypothetical protein C35B8.2 - Caenorhabditis elegans]	2	6.0	97.36	R
8	4.3e-013	-	T.gil14009690[gb AAK51717.1 (AF348175) PA protein [Influenza A virus (A/Hong Kong/1/68(H3N2))]I	9	5.3	82.91	R
9	3.0e-013	-	T.gil14973264[gb AAK75842.1 (AF007469) glycosyl transferase, family 8 [Streptococcus pneumoniae]	9	5.2	94.60	R
10	1.2e-013	-	T.gil13423336[gb AAK23864.1 (AF005862) ribonucleotide reductase-related protein [Caulobacter crescentus]	10	5.1	98.47	R

Figure 4c

ProFound - Search Result Details

Version 4.10.5
The Rockefeller University Edition

Details for rank 1 candidate in search C2BABE89-01DC-7445F3E9

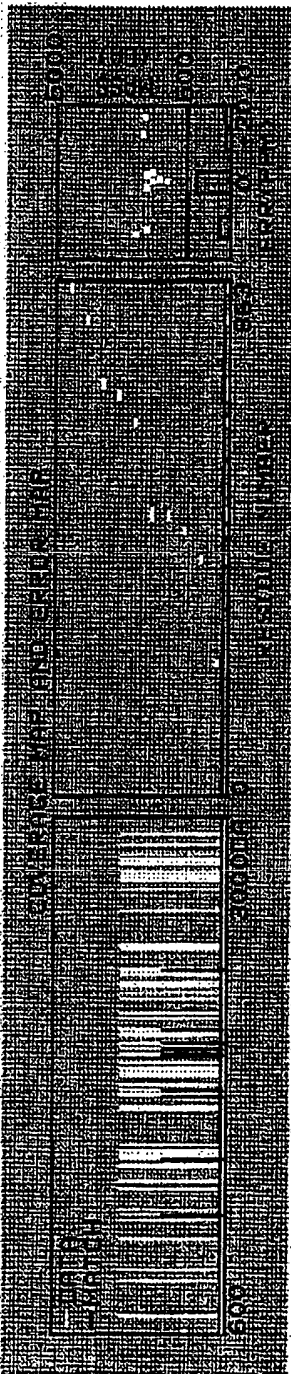
gi5453174|gb|AAD43465.1|AFI13606_1 (AFI13606) UspA1 [Moraxella ostarhialis]

Sample ID : 11020EE20-7-01 [Pass:0]

Measured peptides : 82

Matched peptides : 10

Min. sequence coverage: 18%



Mass (M)	Mono	Computed	Error (ppm)	Start	Residues	Missed	Peptide sequence
1156.584	M	1156.583	0	215	226	0	ANADNAVALGNK
1414.706	M	1414.704	1	437	449	0	NNVEEGLDLSGR
1700.861	M	1700.857	2	463	478	0	ALSNVEEGLDLSGR
1730.868	M	1730.868	0	388	403	0	TLESNVEEGLDLSGR
1894.914	M	1894.916	-1	842	861	0	IGLSWSDAGGLSGVGGSYR
1908.046	M	1908.042	2	787	806	0	QTAGGIANAIATILPSPSR
1926.939	M	1926.964	-13	664	680	0	NQNTLIEQGEALVEQNK
1954.970	M	1954.938	17	681	698	0	AINQELGFAAHADVQDK
2015.980	M	2015.954	13	618	634	0	NIYELAQQQDQHSDDIK
2298.174	M	2298.206	-14	463	483	1	ALSNVEEGLDLSGRLLDQK

Unmatched Monoisotopic Masses:

679.514 684.348 842.510 897.421 1036.535 1126.566 1179.598 1277.702 1300.538 1302.633 1306.630 1307.681
1311.689 1379.718 1383.708 1424.752 1434.769 1438.723 1440.774 1458.736 1474.731 1475.757 1637.850
1638.859 1648.850 1657.807 1700.854 1707.781 1715.823 1716.858 1757.875 1794.809 1807.903 1826.798
1837.968 1851.920 1890.972 1922.930 1940.931 1987.056 1993.976 2064.139 2081.967 2087.974 2150.075
2184.140 2211.104 2221.129 2225.129 2230.218 2239.136 2261.122 2283.164 2297.197 2383.981 2399.049
2408.087 2564.193 2663.346 2705.179 2717.129 2720.281 2748.314 2781.924 2791.004 2792.915 2805.359
2807.327 2825.331 2872.402 2914.526 2921.383

Search again using unmatched masses:

Search for single protein only in All taxa

Figure 4d

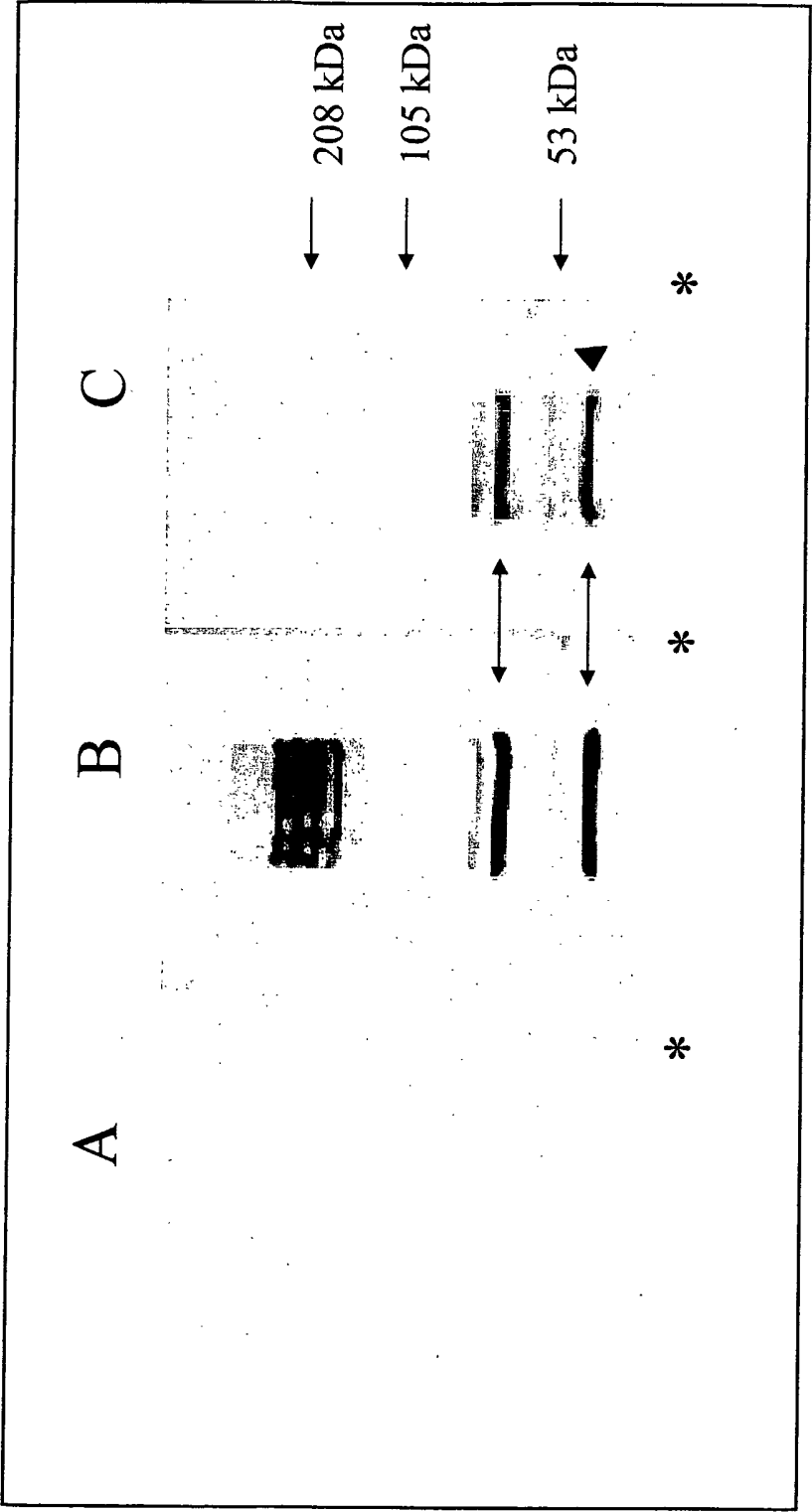


Figure 5

1M¹NKIYKVKKNAAGHLVACSEFAKGHTKKA⁷⁸VLGSLIVGILGMATTASAQKVKGATNKISGGDNNTANGTYLTIGGGDY
79N⁷⁹KTGRYSTIGGG¹⁵⁸LFNEATNEYSTIGSGGYNKAKGRYSTIGGGGYNEATNQYSTIGGGDNNTAKGRYSTIGGGGYNEAT¹⁵⁸
159ENSTVGGGGYNQAKGRNSTVAGGYNNEATGTDSTIAGGRKNQATGKGSFAAGIDNKANADNAVALGNKNTIEGENSV²³⁶A²³⁶
237IGSNNTVKKGQQNVF³¹⁶ILGSNTDIT³¹⁶TNAQNGSVLLGHNTAGKAA³¹⁶TIVNSAEVGG³¹⁶SLTG³¹⁶FAGASKTGN³¹⁶TVSVGKKGKERQ³¹⁶J³¹⁶
317VHV³¹⁷GAGEISDTSTDAVNGSQLHALATVVAQNKADIKDL³¹⁷DDEVGL³¹⁷LGEEIN³¹⁷SLGEIFNNQDAIAKNQADIK³¹⁷TLESNVEEG³¹⁷
397L³⁹⁷LDLSGRLLDQKADIDNNINN³⁹⁷IYELAQ³⁹⁷QDQHSSDI³⁹⁷TKLKN³⁹⁷NVEEGL³⁹⁷LDLSGRLLDQKADL³⁹⁷TKD³⁹⁷IKALESNVEEGL³⁹⁷LDLSGRLL³⁹⁷
481DQKADIAKNQADIAQ⁴⁸¹NQTDIQDLAAYNELQDAYAKQQT⁴⁸¹EADALNKASSANTDRIATAELGIAENKKDAQIAKAQANEN⁴⁸¹
560K⁵⁶⁰DGIAKNQADIQ⁵⁶⁰LHDKKITNLGILHSMVARAVGNNTQGVATNKADIAKNQADIA⁵⁶⁰NNIYELAQ⁵⁶⁰QDQHSSDI⁵⁶⁰KT⁵⁶⁰LAK⁵⁶⁰
639V⁶³⁹SAANTDRIAKNKA⁶³⁹EADASFETLTKNQNTLIEQGEALVEQNKAINQELEGFAAHADVQDKQLQNQADIT⁶³⁹TNKTAIEQNIN⁶³⁹
720RTVANGFEIEKNKAGIATNKQELILQNDRLNRINETNNHQDKIDQLGYALKEQGQHFN⁷²⁰NRISAV⁷²⁰ERTAGGIANAIAIATL⁷²⁰
802PSPSRAGEHHVLF⁸⁰²GS⁸⁰²GYHNGQAAVSLGAAGLSDT⁸⁰²GKSTYKIGLSWSDAGLSGGVGGSYRWK⁸⁰²

Figure 6

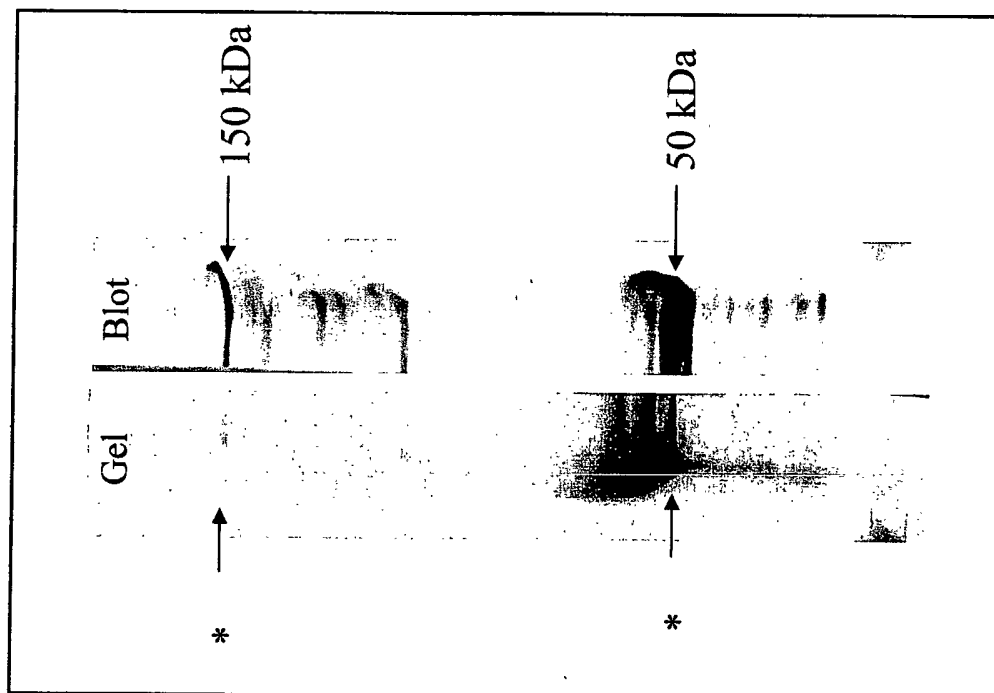


Figure 7

⁴⁶³ **A**LESNVEEGLLDLSGRL⁴⁸⁰

⁴⁸¹ DQKADIAKNQADIAQNQTDIQDLAAYNELQDAYAKQQTAEIDALNKASSANTDRIATAELGIAENKKDAQIAKAQANEN⁵⁵⁹

⁵⁶⁰ KDGIAKNQADIQLDHKKITNLGILHSMVARAVGNNTQGVATNKADIAKNQADIANNIKNIYELAQQQDQHSSDIKTLAK⁶³⁸

⁶³⁹ VSAANTDRIAKNKAEDASFETLTKNQNTLIEQGEALVEQNKAINQELEGFAAHADVQDKQILQNQADITTNKTAIEQNIN⁷¹⁹

⁷²⁰ RTVANGFEIEKNKAGIATNKQELILQNDRNLNRINETNNHQDQKIDQLGYALKEQQGHFNNRISAVERTAGGIANAIAIATL⁸⁰¹

⁸⁰² PPSRAGEHHVLFSGGYHNGQAAVSLGAAGLSDTGKSTYKIGLSWSDAGGLSGGVGGSYRWK⁸⁶³

Figure 8

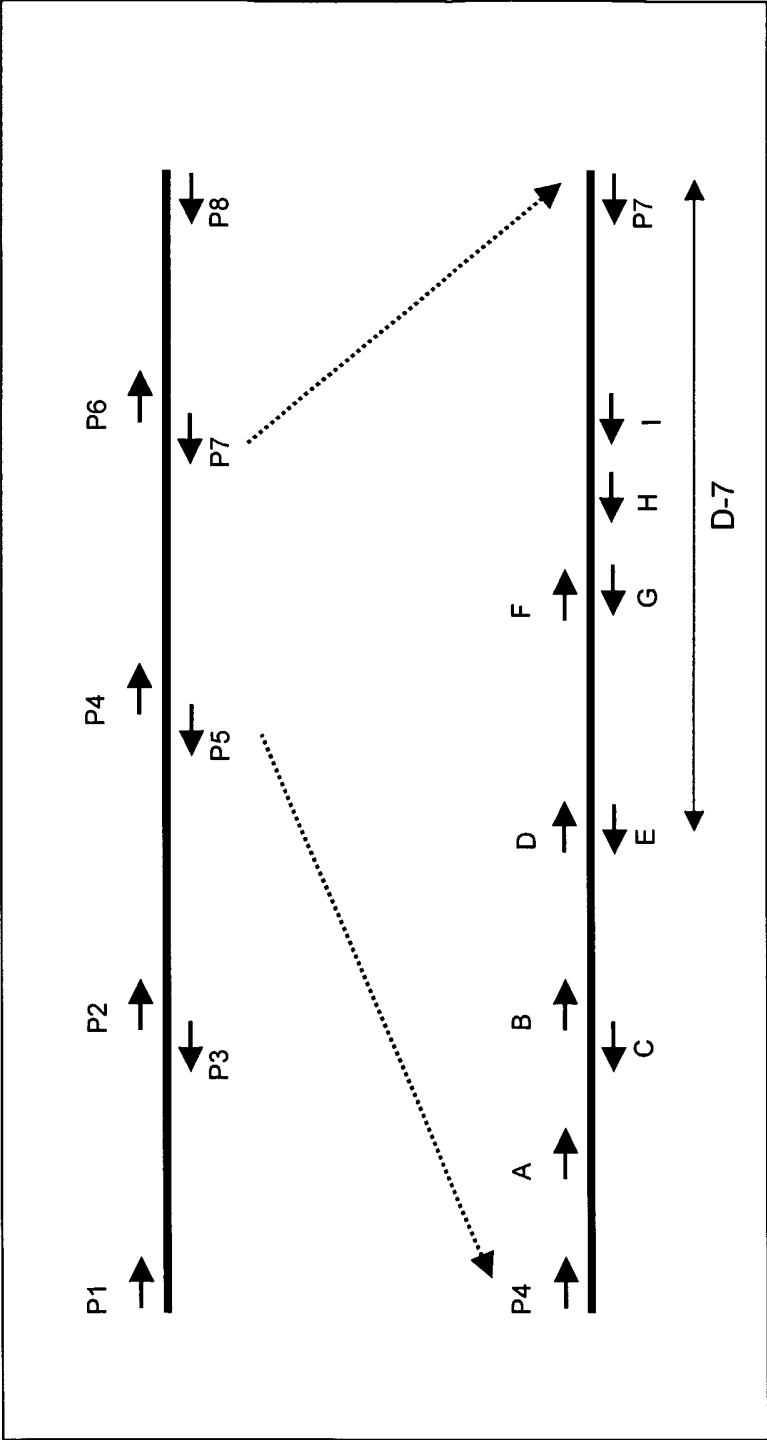


Figure 9

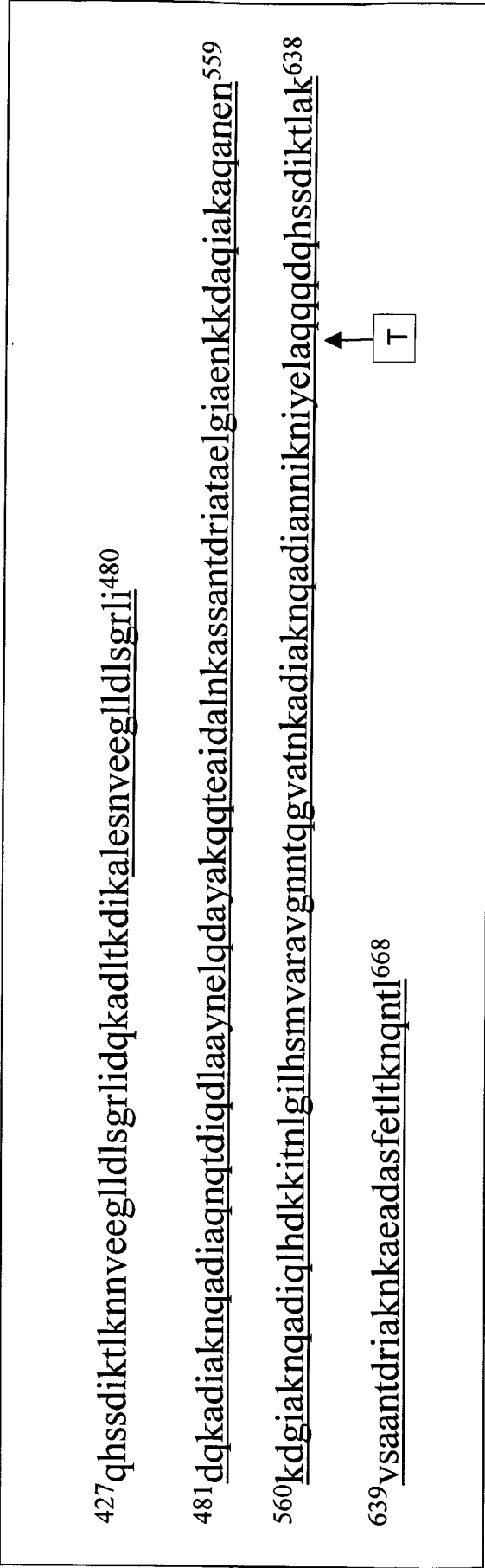


Figure 10

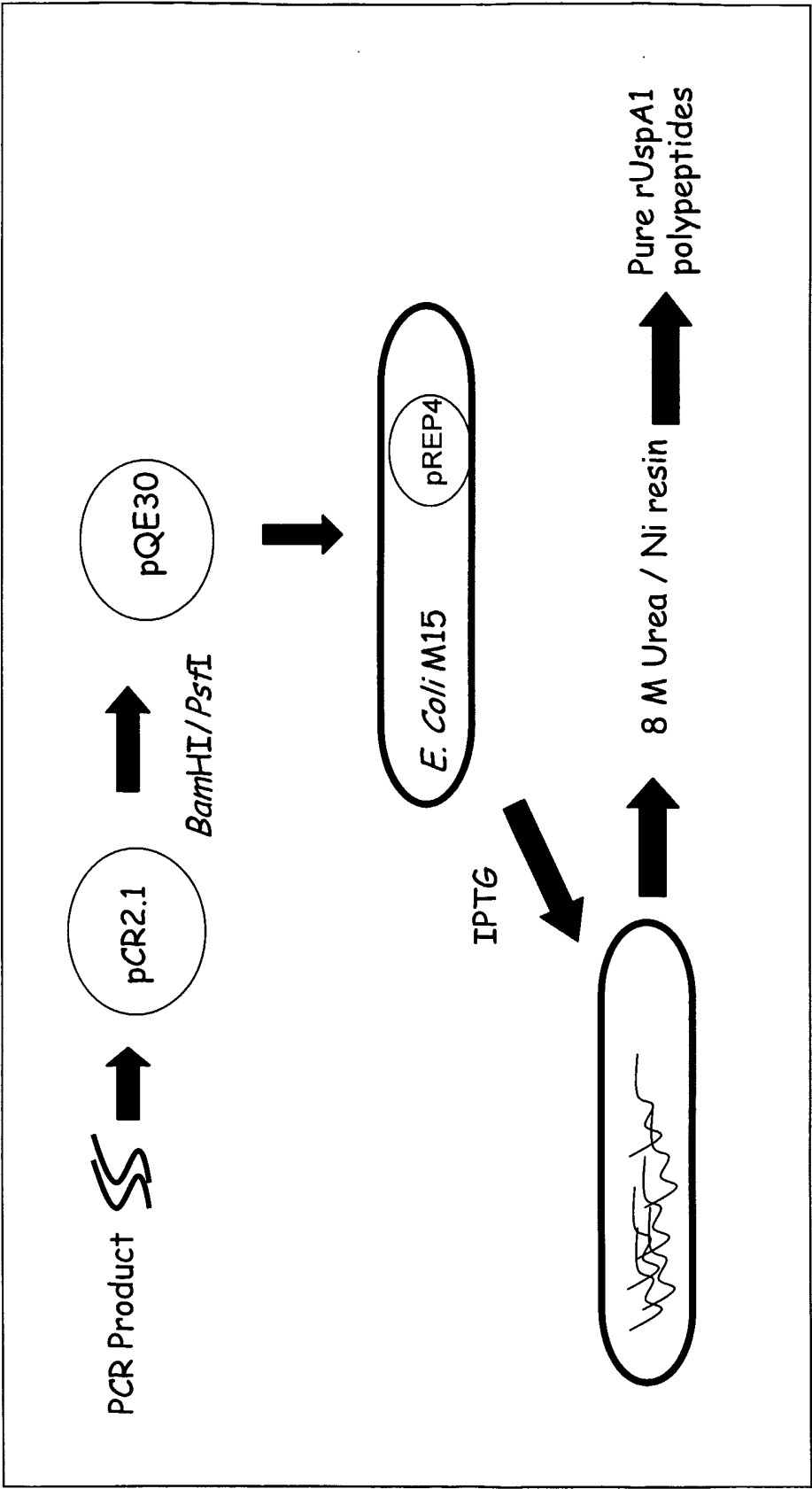


Figure 11

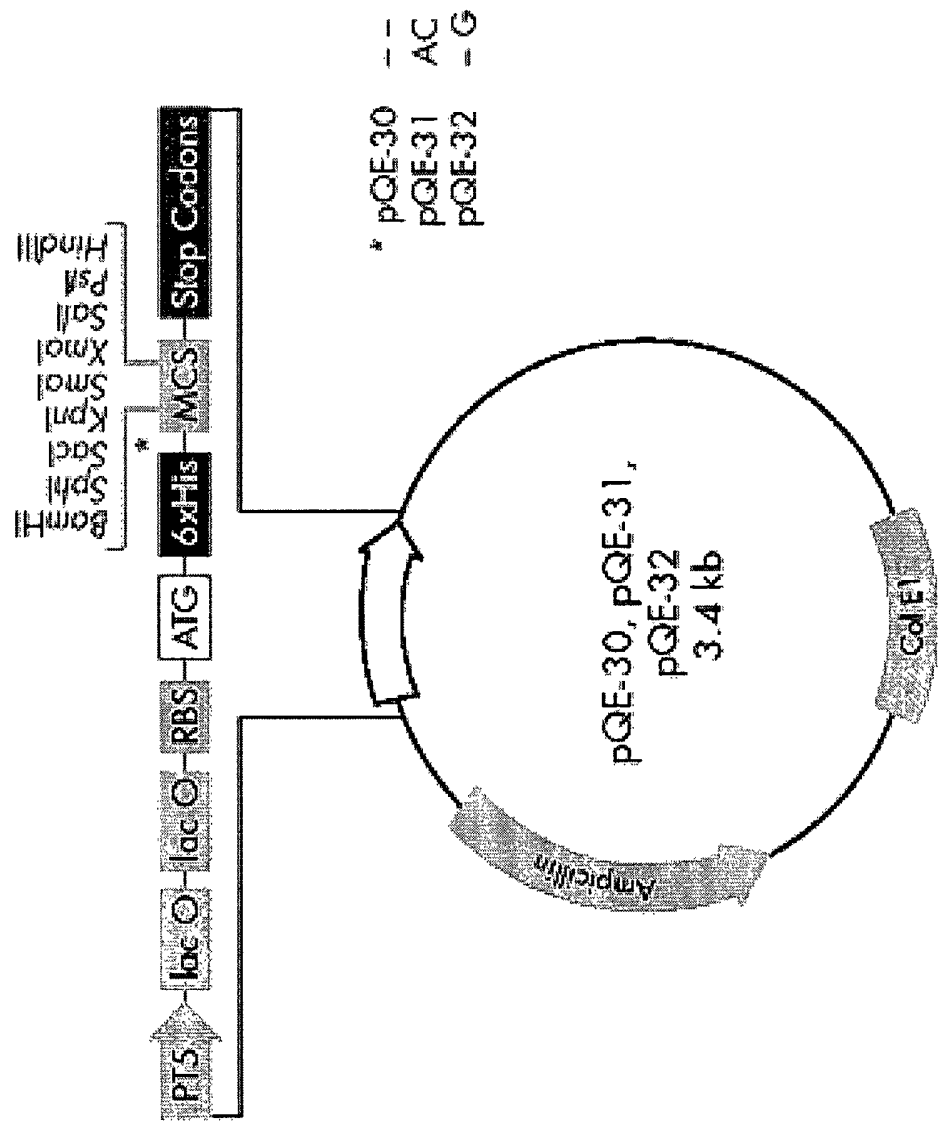


Figure 12

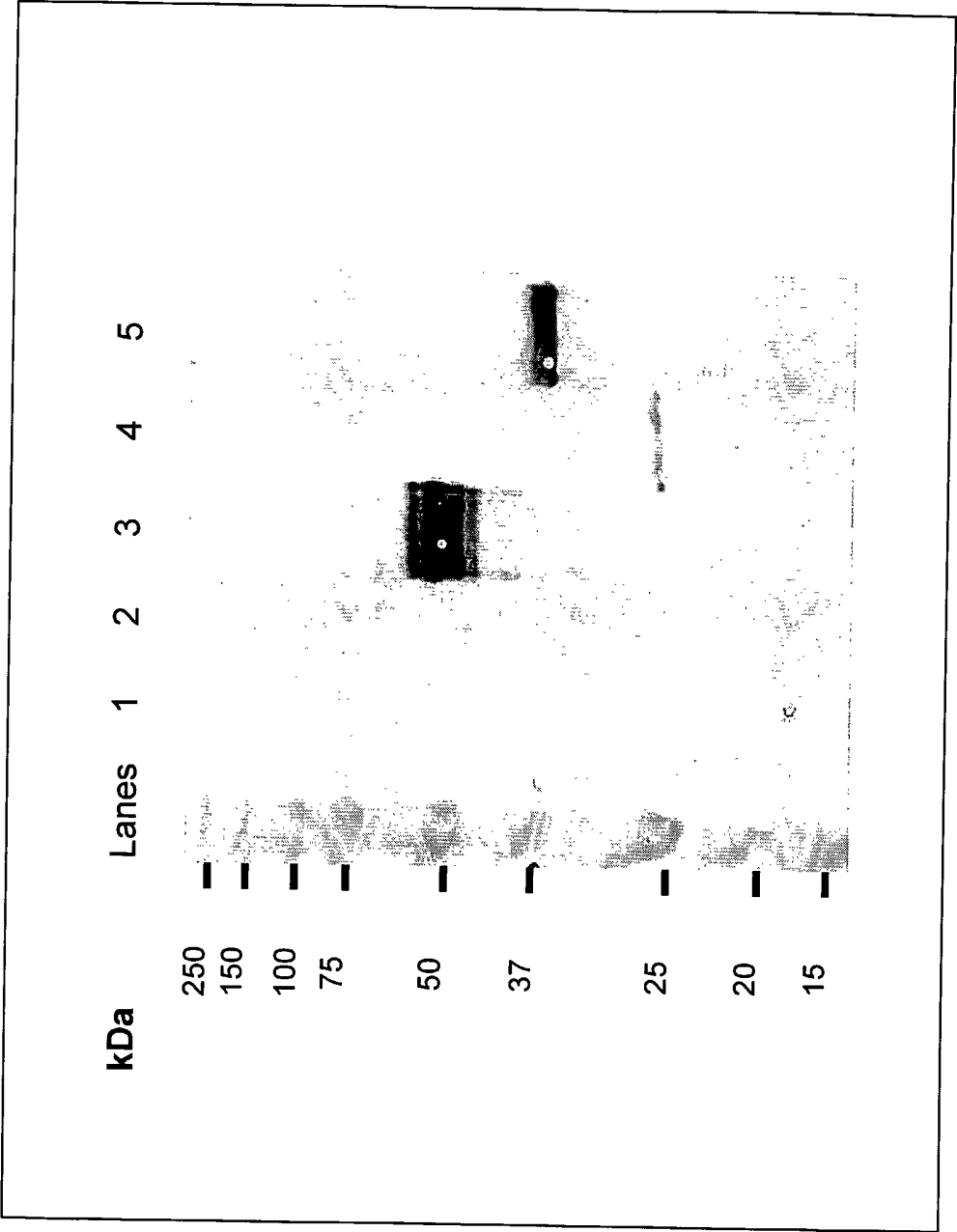


Figure 13

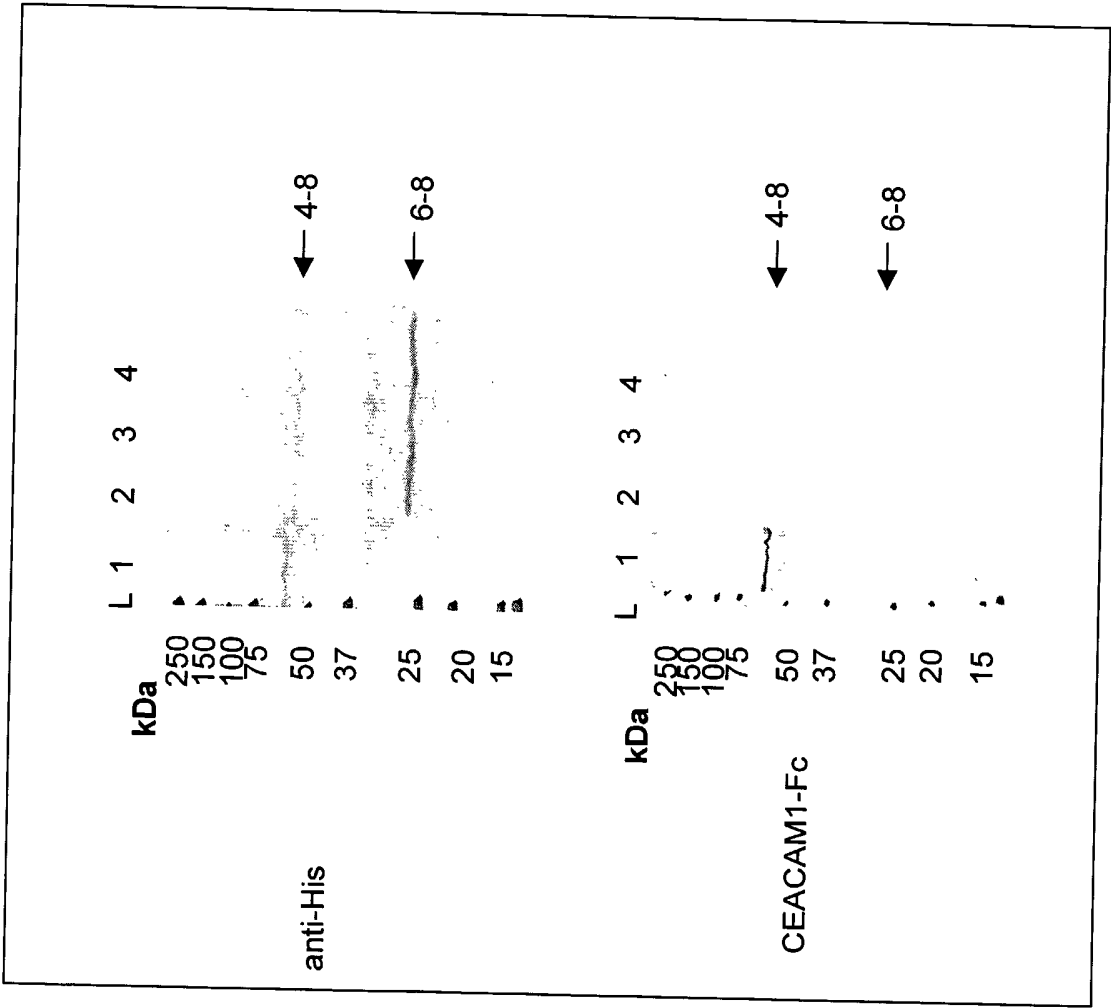


Figure 14

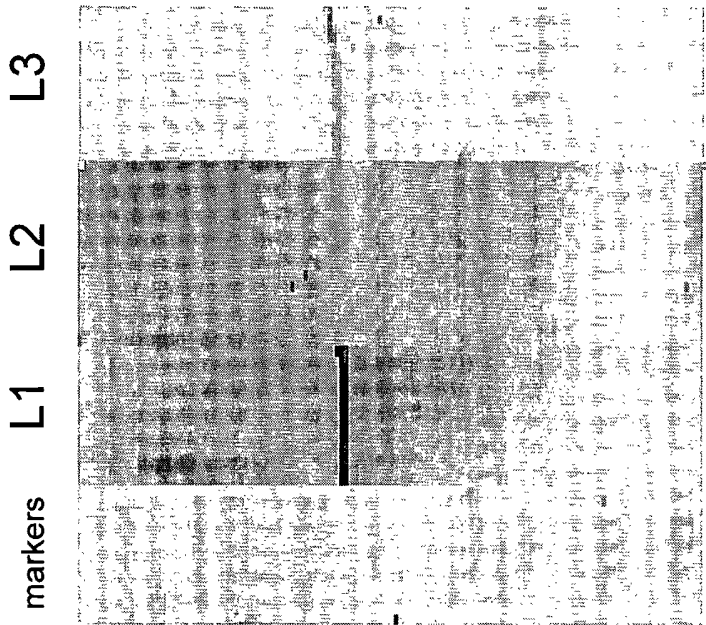
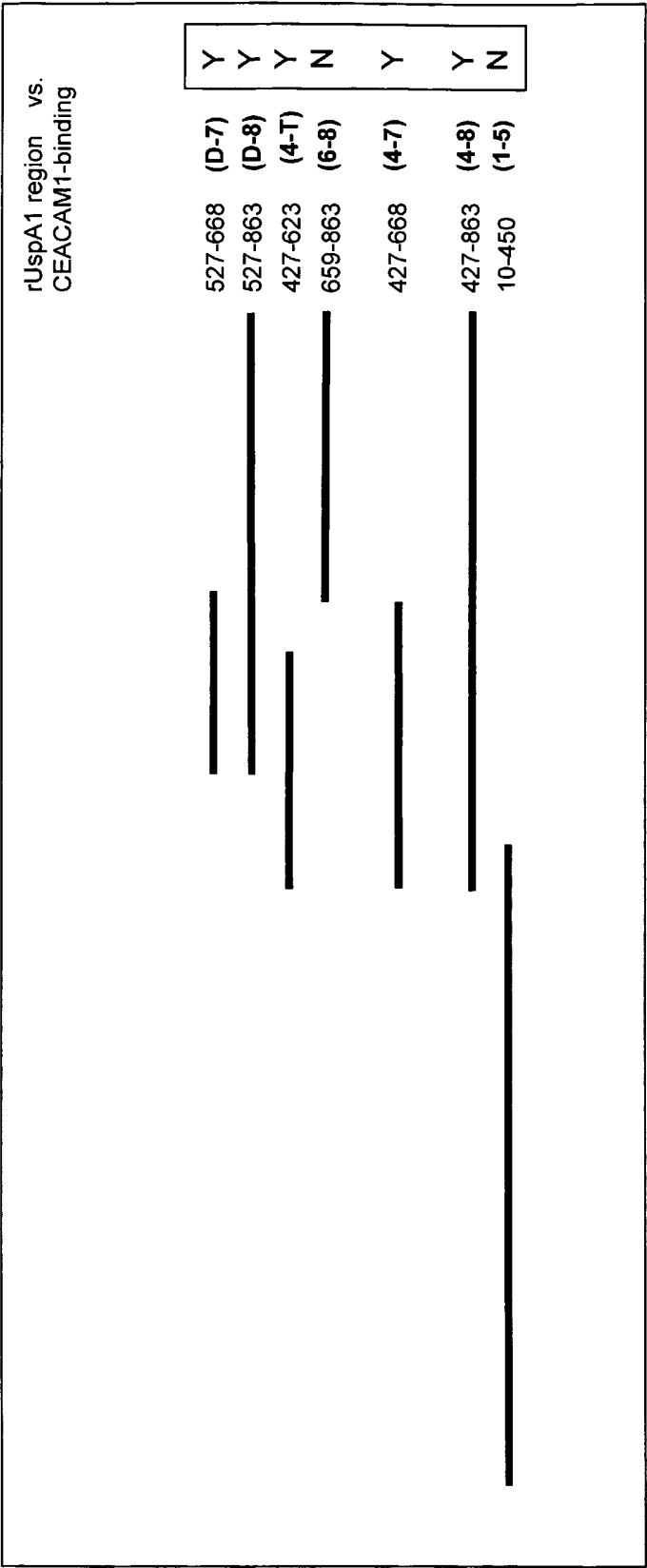


Figure 15



Fragments and predicted MWs (kDa)

D-7	c. 15
D-8	36
4-T	21
6-8	22
4-7	26
4-8	47
1-5	40

for recombinant peptides, add c. 2kDa

Figure 16

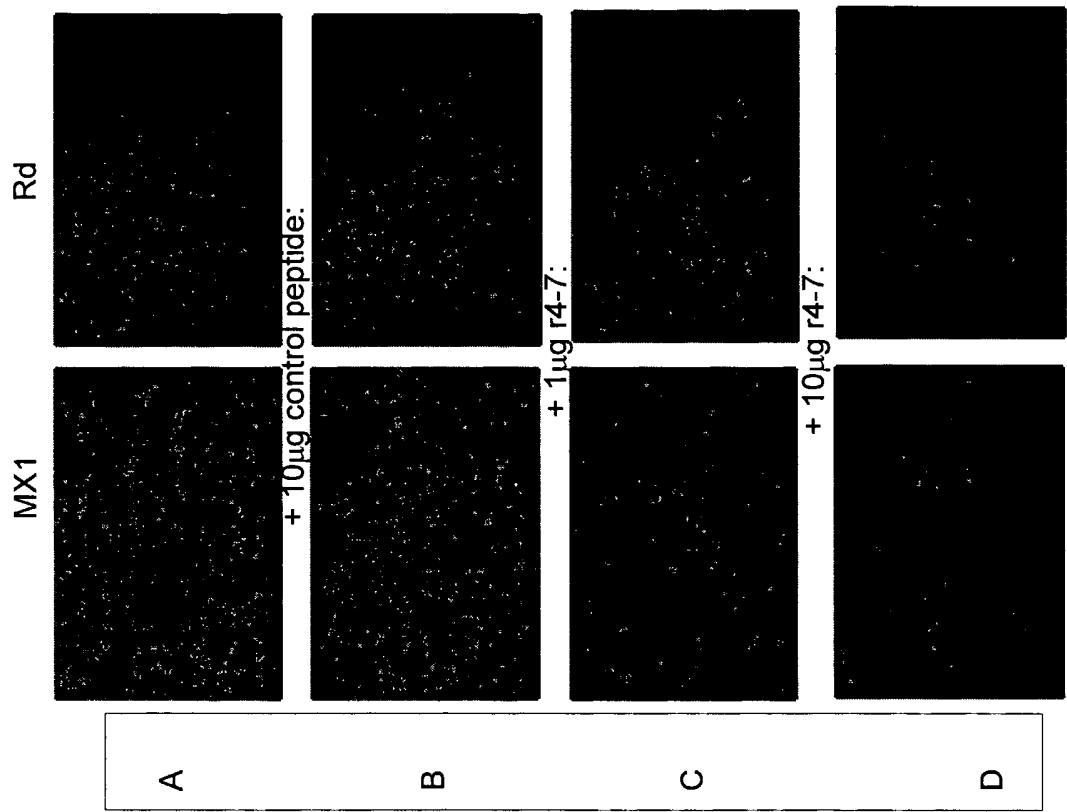


Figure 17