



(43) International Publication Date  
26 May 2016 (26.05.2016)

(10) International Publication Number  
**WO 2016/079755 A1**

(51) International Patent Classification:

C07K 14/34 (2006.01) C12N 15/70 (2006.01)  
C12N 15/10 (2006.01)

(21) International Application Number:

PCT/IN2015/000427

(22) International Filing Date:

17 November 2015 (17.11.2015)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

4045/CHE/2014 20 November 2014 (20.11.2014) IN

(71) Applicant: **BIOLOGICAL E LIMITED** [IN/IN]; 18/1 & 3, Azamabad, Hyderabad - 500020, Andhra Pradesh (IN).

(72) Inventors: **AKSHAY GOEL**; Biological E. Ltd., 18/1 & 3, Azamabad, Hyderabad - 500020, Telangana (IN). **RAVI PRATAP NARAYAN MISHRA**; Biological E. Ltd., 18/1 & 3, Azamabad, Hyderabad - 500020, Telangana (IN). **NARENDER DEV MANTENA**; Biological E. Ltd., 18/1 & 3, Azamabad, Hyderabad - 500020, Telangana (IN). **MAHIMA DATLA**; Biological E. Ltd., 18/1 & 3, Azamabad, Hyderabad - 500020, Telangana (IN).

(74) Agent: **KAYSER, Avesh**; Kayser & Company, 2nd floor, 20, Raja Bahadur Mansion, Ambalal Doshi Marg, Fort, Mumbai - 400001, Maharashtra (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— of inventorship (Rule 4.17(iv))

**Published:**

— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

— with sequence listing part of description (Rule 5.2(a))

(54) Title: CODON OPTIMIZED POLYNUCLEOTIDE FOR HIGH LEVEL EXPRESSION OF CRM<sub>197</sub>

(57) Abstract: The present invention relates to high level expression of bacterial toxoid or toxin protein of pharmacological interest by means of an optimized novel polynucleotide sequence and host transformed with the said polynucleotide. Specifically, the invention provides a method for high production of polypeptide CRM<sub>197</sub> wherein, the polynucleotide of the invention is used to transform a suitable host resulting in over-expression of corresponding proteins and a method for isolating the expressed polypeptide. More particularly, the present invention relates to high level expression of CRM<sub>197</sub> in *Escherichia coli* and a method for the isolation and purification thereof.



WO 2016/079755 A1

## CODON OPTIMIZED POLYNUCLEOTIDE FOR HIGH LEVEL EXPRESSION OF CRM<sub>197</sub>

### Field of the Invention

The present invention relates to a high level expression of bacterial toxoid by means of an optimized novel polynucleotide sequence and host transformed with the said polynucleotide.

The invention also provides a method for high production of polypeptide CRM<sub>197</sub> wherein, the polynucleotide of the present invention is used to transform a suitable host resulting in over-expression of corresponding proteins and a method for isolating the expressed polypeptide.

### Background of the Invention

Diphtheria toxin (DT) is a protein exotoxin that is synthesized and secreted by *Corynebacterium diphtheriae*. The toxigenic strains of *Corynebacterium diphtheriae* contain a bacteriophage lysogen carrying the toxin gene. Mature form of DT is synthesized as a 535 amino-acid containing single polypeptide, which is derived from an initial 536 pro-peptide which undergoes proteolysis at positions 190, 192 and 193 to form the mature toxin. This splicing or proteolysis results into two subunits, A and B which are joined together by a disulfide bridge (*Moskang et al Biol. Chem.* 264: 15709-15713, 1989). Subunit A is catalytically active NAD-dependent ADP-ribosyl transferase portion. It is responsible for rendering Elongation Factor-2 (EF-2) inactive, and hence down regulates the protein synthesis in a target cell.

Diphtheria toxin is highly cytotoxic; a single molecule can be lethal to a cell, and a dose of 10 ng/kg can kill animals and humans. In course of providing an artificial immunity by injection of this toxoid, a process of detoxification should be included to render it safe for consumption by the recipient. Conventionally it was detoxified by chemical modification of the natural forms of DT, in such a manner that it still retains the required antigenicity required in a vaccine preparation.

Subsequently, a genetically detoxified form of Diphtheria Toxin known as Cross Reacting Material 197 or CRM<sub>197</sub> was introduced; which essentially retains the immunological cross-reacting properties of DT.

CRM<sub>197</sub> has been used in preparation of conjugate vaccines including *Corynebacterium diphtheria*, *Hepatitis B*, *Bordetella pertussis*, *Clostridium tetani*, *Neisseria meningitides*, *Streptococcus pneumonia*, *Haemophilus influenza*. It was generated by nitrosoguanidine mutagenesis of the toxigenic corynephage  $\beta$ , which was then used to infect *Corynebacterium diphtheria*. (Uchida et al *Nature New Biology* (1971) 233; 8-11, *Nucleic Acids Res.* 1984 May 25; 12(10):4063-9)

CRM<sub>197</sub> has been studied for its potential use as a DT booster or vaccine antigen. The CRM<sub>197</sub> protein has the same molecular weight as DT; but differs in a single base change in the A subunit i.e. a base change in the polynucleotide sequence of wild type DT, wherein a replacement of Guanine to Adenine results into an amino acid substitution at position 52, resulting into glutamic acid in CRM<sub>197</sub> instead of glycine (Giannini G. et al., 1984). This point mutation results in a significant loss of toxicity and renders CRM<sub>197</sub> safe for human use.

Production of significant quantities of diphtheria toxins such as CRM<sub>197</sub> for use in vaccines has been hindered due to low level of expression in wild type bacteria. This problem has been addressed previously by expressing CRM<sub>197</sub> in *Escherichia coli* by Bishai et al., (*J. Baeteriol.* 189:5140-5151), who describe the expression of a recombinant fusion protein containing diphtheria toxin (including the tox signal sequence); but this led to the production of degraded protein. The low yield in active form is also associated with degradation, improper folding, or both, depending on the specific characteristics, e.g., size and secondary structure, of the toxin. Hence as with most biopharmaceuticals, there is additional loss of expressed protein that occurs during the purification steps of CRM<sub>197</sub>, whereby maintaining a biologically active form of CRM<sub>197</sub> poses a challenge. Therefore, there is need to achieve high level expression of bacterial toxoid CRM<sub>197</sub> in an active form.

WO 2011/042516 discloses an improved process for making a bacterial toxin by periplasmic expression comprising the steps of a) growing a culture of the bacterial host cell containing an expression vector in which particular signal sequence are linked to the sequence of a bacterial toxin and b) inducing expression of the polypeptide containing particular signal sequence linked to a bacterial toxin such that a bacterial toxin is expressed periplasmically.

WO 2013/178974 A1 discloses a process for the intracellular expression of CRM<sub>197</sub> in an *Escherichia coli* host, comprising expressing a vector comprising a gene encoding CRM<sub>197</sub> operably linked to a Promoter and at least one perfect Palindrome Operator sequence.

WO 2015/134402A1 discloses a process for producing a recombinant CRM<sub>197</sub> in a reduced genome of *Escherichia coli* host comprising incubating a reduced genome *Escherichia coli* comprising an expression vector comprising a nucleotide sequence encoding a CRM<sub>197</sub> protein fused to a signal sequence that directs transfer of the CRM<sub>197</sub> protein to the periplasm operably linked to an expression control sequence under conditions suitable for the expression of the recombinant CRM<sub>197</sub> protein, whereby a yield of at least 1 gram per liter of soluble CRM<sub>197</sub> is obtained and wherein the native parent *Escherichia coli* strain is a 12 strain, preferably K12 MG1655.

US 2012/0128727 A1 discloses an isolated nucleic acid molecule which encodes polypeptide CRM<sub>197</sub>, an expression vector comprising the isolated nucleic acid molecule and a method for recombinant production of a CRM<sub>197</sub> tag fusion protein, comprising culturing the recombinant cell under conditions favoring production of said CRM<sub>197</sub> tag fusion protein, and isolating said fusion protein.

US 2012/0289688 A1 discloses a process for periplasmic expression of a recombinant polypeptide by (A) Growing a culture of a gram-negative host cell; and (B) Inducing expression of a polypeptide such that a protein is expressed periplasmically; wherein one or more of the following steps is actioned during expression: (i) The pH of step a)

is lower than the pH of step b); (ii). The temperature of step a) is higher than the temperature of step b); or (iii). The substrate feed rate of step a) is higher than the substrate feed rate of step b).

US 2014/0050758 A1 discloses a process for periplasmic expression of a bacterial toxoid comprising the steps of: a) growing a culture of a gram negative host cell in a fermentation medium, wherein the host cell is transformed with a polynucleotide, and wherein the polynucleotide encodes the bacterial toxoid and a periplasmic signal sequence; inducing expression of the bacterial toxoid;

b) maturing the host cell, wherein the maturing step comprises: I) subjecting the host cell to a pH shock: II) incubating the host cell with no feed addition; or III) subjecting the host cell to a temperature below -20°C; and

c) extracting the bacterial toxoid from the host cell wherein the extraction process comprises osmotic shock wherein the gram negative host cell is selected from the group consisting of *Escherichia coli*, *Pseudomonas* and *Moraxella*, wherein the host cell is alive during step b) and wherein the process is carried out in a fermenter which contains 10-5000 litres of culture.

US 8,530,171 discloses a method for producing a recombinant toxin protein in a *Pseudomonas* host cell, said method comprising: ligating into an expression vector a nucleotide sequence encoding the toxin protein; transforming the *Pseudomonas* host cell with the expression vector; and culturing the transformed *Pseudomonas* host cell in a culture media suitable for the expression of the recombinant toxin protein; wherein the recombinant carrier protein is CRM<sub>197</sub>, and wherein the recombinant protein is produced at a yield of soluble or active CRM<sub>197</sub> protein of about 0.2 grams per liter to about 12 grams per liter.

Conjugated polysaccharide vaccines that use CRM<sub>197</sub> as a carrier protein have been approved for human use. These include: Menveo® (Novartis Vaccines and Diagnostics), a vaccine indicated for preventing invasive meningococcal disease caused by *Neisseria meningitidis* subgroups A, C, Y, and W-135; Menjugate® (Novartis Vaccines), a

meningococcal group C conjugate vaccine; and Prevnar® (Wyeth Pharmaceuticals, Inc.), a childhood pneumonia vaccine that targets thirteen serotypes of *Streptococcus pneumoniae*, and HibTITER® (Wyeth), a *Haemophilus influenzae* type b vaccine. In addition, CRM<sub>197</sub> has potential use as a boosting antigen for *C. diphtheria* vaccination and is being investigated as a carrier protein for use in other vaccines.

There has recently been a growing interest in CRM<sub>197</sub> because of its potential antitumor action relating to its capacity to bind the soluble form of HB-EGF (*Mekada et al, US Patent Publication NO. 2006/0270600A1*). This antitumor function is attributable not only to CRM<sub>197</sub>, but also to other non-toxic derivatives of the DT toxin (e.g. the double mutant DT52E148K, or the fusion protein GST-DT). These mutants have been constructed by PCR, starting from the gene encoding CRM<sub>197</sub>. In said studies, however, the whole CRM<sub>197</sub> was produced using cultures of *C. diphtheria*, grown at 35° C for 16-17 hours. The CRM<sub>197</sub> was purified from the supernatant by means of an initial precipitation with ammonium sulphate, followed by three successive steps in ion exchange and hydrophobic chromatography (*Mekada et al.*).

Hence, there is an evident need for an alternative method for the production of CRM<sub>197</sub> with high yield and cost-effective manner. Therefore, a method for economically producing CRM<sub>197</sub> would greatly facilitate vaccine research, development and manufacturing.

### **Objective of the Invention**

The main objective of the present intention is to provide an optimized polynucleotide for high level expression of CRM<sub>197</sub>.

Yet another objective is to provide a tunable process for controlling the expression of polypeptide as to obtain CRM<sub>197</sub>.

Yet another objective is to provide a high level expression process for commercial production of CRM<sub>197</sub> in pure form with high yield.

**Summary of the Invention**

The present invention provides an optimized polynucleotide sequence comprising of SEQ ID NO. 2 and its variants which are at least 70% homologous to the said optimized polynucleotide sequence SEQ ID NO. 2.

In another embodiment, the present invention provides an optimized polynucleotide sequence (SEQ ID NO. 2) and its structural variants selected from but not limited to SEQ ID NO. 3, 4, 5, 6, 7, 8, 9 and 10 useful for high level expression of polypeptide CRM<sub>197</sub>.

In yet another embodiment, the present invention provides an optimized polynucleotide sequence (SEQ ID NO. 2) and its variants like SEQ ID NO. 3, 4, 5, 6, 7, 8, 9 and 10 which are at least 70 to 88 % homologous to the said optimized polynucleotide sequence SEQ ID NO. 2.

The present invention further provides a process for the production of polypeptide, comprising steps of:

- a) selecting an optimized polynucleotide sequence essentially consisting of SEQ ID NO. 2 or its variants which are at least 70% homologous to SEQ ID NO.2,
- b) optionally ligating the polynucleotide sequences of step (a) into a suitable vector,
- c) inserting or transforming the polynucleotide sequence into *Escherichia coli* host cell,
- d) culturing the transformed host cell in a culture media for high level expression of the polypeptide,
- e) maintaining the induction temperature between 10 to 40 °C to produce polypeptide,
- f) extracting the bacterial polypeptide from the host cell, followed by purification to obtain pure polypeptide with high yields.

The polypeptide obtained above is suitably used as a carrier protein for preparation of conjugated immunogenic preparations.

### **Brief Description of the Drawings**

Figure 1: illustrates SDS-PAGE electrophoretic gel run corresponding to the expressed CRM<sub>197</sub> which is encoded by polynucleotide SEQ ID NO. 2; where Lane 1: Standard Molecular Mass Marker; Lane 2 and 3: CRM<sub>197</sub> which was separated and purified from the total cellular proteins of *E. coli* culture extracts and run on non-reducing and reducing SDS-PAGE, respectively.

Figure 2: illustrates western blot analysis of purified CRM<sub>197</sub> using rabbit polyclonal antibodies; Lane 1: Molecular weight ladder. Lane: 2 and 3 includes CRM<sub>197</sub> samples electrophoresed under reducing and non-reducing conditions, respectively.

Figure 3: SDS PAGE shows purified soluble fraction Lane 1: Reference protein; Lane 2: protein molecular weight marker; Lane 3-10: Pooled Polypeptide CRM<sub>197</sub>.

Figure 4: Electrophoretic gel run (SDS PAGE 12%) showing test conducted on solubilization of CRM<sub>197</sub>, Lane 1: Urea solubilized fraction of CRM 197; Lane 2: Supernatant; Lane 3: Sample from first pellet wash; Lane 4: Sample pooled from 2<sup>nd</sup> pellet wash.

Figure 5: Size exclusion chromatography (SEC-HPLC) wherein major eluted peak shows the presence of CRM<sub>197</sub> in the sample.

Figure 6: Peptide mass fingerprint (mass spectrometry) of polypeptide CRM<sub>197</sub> to define primary amino acid sequence identity. Recombinant CRM (BioE rCRM) of the present invention had 100% sequence similarity with the reference CRM<sub>197</sub> sequence.

Figure 7: N-Terminal sequence confirmation of rCRM<sub>197</sub> by Edman degradation. The 10 amino acid sequence GADDVVDSSK (N-Term acetyl) shows starting portion of purified polypeptide. The first amino acid is identified as G.

Figure 8: CD spectra of the recombinant CRM (BioE rCRM) of the present invention was overlapped with the reference CRM<sub>197</sub>. The secondary structure parameter were also analysed and showed the similarity with reference. The result shows that recombinant CRM (BioE rCRM) of the present invention is structurally similar to the reference CRM<sub>197</sub>.

Figure 9: Confirmation of structural equivalence of recombinant CRM (BioE rCRM) of the present invention with reference using fluorescence assay. Overlay DSF profile of recombinant CRM (BioE rCRM) of the present invention with reference CRM<sub>197</sub> (C7 CRM). The data confirms the similarity of recombinant CRM (BioE rCRM) of the present invention with the reference C7- CRM<sub>197</sub>.

Figure 10: Confirmation of disulphide bonds. The recombinant CRM (BioE rCRM) of the present invention was analysed for the presence of correct disulphide bonds in the protein. It is confirmed that two disulphide bonds present in the protein first links amino acid 186 to 201 and second bond links amino acid 461 to 471. The mass spectrometry method was used to analyse the disulphide linkages in CRM<sub>197</sub>.

Figure 11: Confirmation of antigenic similarity of recombinant CRM (BioE rCRM) of the present invention with reference CRM<sub>197</sub> by CRM<sub>197</sub> specific ELISA. All the CRMs coming from difference source showed similar recognition profile with monoclonal antibodies.

### **Detailed Description of the Invention**

The polypeptide expressions for use as pharmaceutical product or vaccines requires achieving high biomass and/or productivity of the host cell line. The efficiency of polypeptide production can be significantly diminished in absence of multiple factors,

which includes use of an optimal polynucleotide sequence encoding that polypeptide. The genetic code is known to exhibit degeneracy, which amounts to the variance in the polynucleotide sequence encoding the same amino acid sequence. The rate of synthesis of amino acid chain is a determinant factor in the overall expression levels from an individual gene, which effect the design of the expression construct, is of high significance. Thus the construction of an optimal polynucleotide sequence is important in determining the overall expression levels of a polypeptide and has to be well regulated. It includes, but is not limited to the frequency with which the codons are preferred in an organism or in case of artificial vehicles or vectors, the nearest frequency desired. This in turn reflects tRNA abundance or the cognate cellular tRNA frequencies from which the synonymous codon choice patterns has to be carefully selected. Additional factors also include the potential for formation of secondary structures, mRNA levels and RNA stability, subsequent intended manipulations to be carried out, synthesis routes and so on. The occurrence of these structures has to be carefully regulated as the choice of these patterns differs with the optimizations for individual protein of interest and expression hosts.

Accordingly, the main embodiment of the present invention provides an optimized polynucleotide sequence (SEQ ID NO. 2) and its structural variants.

In another embodiment, the invention provides an optimized polynucleotide sequence (SEQ ID NO. 2) and its structural variants selected from but not limited to SEQ ID NO. 3, 4, 5, 6, 7, 8, 9 and 10 having equal to or more than 70% similarity useful for high level expression of polypeptide for CRM<sub>197</sub>.

Periplasmic expression refers to the secretion of the expressed product from the intended gene of interest (such as a bacterial toxoid or Diphtheria Toxoid) in the periplasmic space within a host cell.

Cytoplasmic expression refers to the expression of protein in the cytoplasmic compartment of the cell, enclosed within cell membrane.

Induction of expression refers to the step performed to induce the expression from the polynucleotide so that the product is obtained at an accelerated rate, this may involve addition of suitable inducing agent such as IPTG, arabinose, maltose and the like.

The optimized sequence of the present invention is applicable to other variants of SEQ ID NO. 2 selected from but not limited to SEQ ID NO. 3, 4, 5, 6, 7, 8, 9, 10 and also include sequences in the production of derivatives of SEQ ID NO. 1, wild type Diphtheria toxin which retains the same inflammatory and immunostimulatory properties and is capable of binding to the cell receptor HB-EGF, but differs from CRM<sub>197</sub> in a single amino acid substitution and lack of cellular toxicity on target host.

The polynucleotide sequence of the CRM<sub>197</sub> may be derived from the sequence of Diphtheria Toxin (Greenfield, L. et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:6853-6857), or by using the amino acid sequence of CRM<sub>197</sub> given by *Giannini G. et al., (1984)* as reference. The wild type polynucleotide sequence thus obtained was optimized for high expression in the host cell, more preferably Gram negative cell, more preferably *Escherichia coli* as host cell. Such a polynucleotide sequence of the present invention can be prepared by chemical synthesis or by means of an assembly procedure.

In yet another embodiment, there is provided a process for the intracellular expression of CRM<sub>197</sub> in a host cell wherein an expression construct with regulatory sequence provides for the expression of the polynucleotide of the invention. This polynucleotide sequence may be associated with a signal sequence for directed transport of the encoded polypeptide. It may be operably linked to periplasmic signal sequence which provides the expression targeted to be secreted in the periplasmic space of the host.

The present invention also provides high level production of CRM<sub>197</sub>, wherein periplasmic expression is brought about by providing a suitable induction temperature for the expression of polynucleotide, without any heterologous sequence for directed transport into the periplasmic space.

Optionally the polynucleotide of the invention may also be associated with polynucleotides of tag polypeptides. The presence of a tag is also known to enhance the stability and solubility of the protein in the cytoplasm and for its subsequent purification.

These tags can be associated at 5' terminus or 3' terminus, singly or in combination pertinent to multi-tagging, with an oligonucleotide sequence that encodes a tag polypeptide to facilitate its cytoplasmic stability and/or subsequent purification using matrices and resins with a high affinity for the various tag peptides. Various tags which can be used according to the invention include HA Tags (haemagglutinin), MYC Tag, Strep II, FLAG, HPC (heavy chain of protein C), glutathione-S-transferase (GST), maltose-binding protein (MBP), cellulose-binding protein (CBD) and chitin-binding protein (CBP).

The polynucleotide of invention may also be incorporated in a vector construct comprising regulatory sequence, with molecular techniques well known in art (See *Sambrook et al., Molecular Cloning, 2nd ed., (1989)*). This includes but is not limited to, a suitable promoter, origin of replication, ribosomal binding site, transcription termination sequence, selectable markers and multiple cloning site. In particular, a plasmid with an efficient and specific construct is preferred; such as one including T7 Promoter specific for RNA polymerase enzyme of the phage T7. Such methods may be referred to but are not limited to one disclosed in U.S. Patent Application NO. 2012/0128727; U.S. Pat. NO. 5,055,294; U.S. Pat. NO. 5,128,130; U.S. Pat. NO. 5,281,532; U.S. Pat. NO. 4,695,455; U.S. Pat. NO. 4,861,595; U.S. Pat. NO. 4,755,465 and U.S. Pat. NO. 5,169,760. A plasmid system for producing CRM<sub>197</sub> protein in *Corynebacterium diphtheriae* is also described in U.S. Pat. NO. 5,614,382.

In one embodiment, the host cell is a gram negative host cell. Host cell expression systems like *Escherichia coli*, *Bacillus sp.*, *Pseudomonas sp.*, have been extensively discussed in the production of proteins. In one embodiment the polynucleotide of the invention is preferably used for the intracellular expression of CRM<sub>197</sub> in *Escherichia*

*coli* wherein the host strain is selected from BL21 (DE3), BL21 A1, HMS174 (DE3), DH5 $\alpha$ , W3110, B834, origami, Rosetta, NovaBlue (DE3), Lemo21 (DE3), T7, ER2566 and C43 (DE3).

In a preferred embodiment, the present invention provides a polynucleotide sequence (SEQ ID NO. 2) encoding bacterial toxoid which is optionally ligated into a vector, followed by its insertion in a host cell. The insertion into host cell may be performed by any of the methods known in the art. Such an insertion or transformation may be performed by a physical or a chemical method of transformation. Subsequently, the converted colonies are selected on petri dishes with added antibiotic.

Suitable vectors used in the present invention include but not limited to pET9a, pET3a, pET3b, pET3c, pET5a, pET5b, pET5c, pET9b, pET9c, pET12a, pTWIN1, pTWIN2, pET12b, pET12c, pET17b and in general, all the vectors that have a strong phage T7 promoter (e.g. pRSETA, B and C [Invitrogen]) and pTYB1, pTYB2, pTYB3 and pTYB4.

In another embodiment, *Escherichia coli* cells are used to express the polynucleotide encoding CRM<sub>197</sub>. The inserted Polynucleotide is verified for proper orientation and position by sequencing. The resultant construct is used to transform host cells by any of the known chemical or physical methods. For example electroporating host cells with an electric field in range 6.5 kV.cm<sup>-1</sup> to 25kV.cm<sup>-1</sup>, a preferred chemical method here which is used to transform host. These cells are allowed to grow for 30 minutes to 120 minutes at 25 to 40 °C in a suitable medium as LB or SOC medium and then transferred to selection media petriplates for 10 to 36 hours, 25 to 40 °C where the positive colonies containing the polynucleotides of invention are selected.

The selection of positive colonies can be done with or without markers. Suitable markers which can be used are selected from, but not limited to, antibiotics such as ampicillin, kanamycin and the like.

The polynucleotide encoding the full length CRM<sub>197</sub> protein is cloned adjacent to T7 *lacI* promoter that drives the expression of protein in T7 polymerase positive strains of *Escherichia coli*. The expression of polynucleotide is stringently controlled by T7 promoter which is induced in the presence of IPTG or in auto-induction mediums

The parameters for culturing the host are optimised for high level expression of CRM<sub>197</sub> protein. In one embodiment, the culture media components, culture conditions including growth temperature, concentration of inducers and induction time is optimized. The culture media used may be selected from, but not limited to, chemically defined media, LB (Luria-Bertani), TB (Terrific Broth), SOB (*Super Optimal Broth*), SOC (Super Optimal broth with catabolic repressor), YT broth (Yeast Extract and Tryptone), Super broth, rich media, minimal media, mineral media and the like. The ingredients of media includes, but is not limited to, a carbon source such as, e.g., glucose, sucrose, or glycerol, organic nitrogen source, such as peptone, tryptone, amino acids, or a yeast extract, inorganic nitrogen source is used and this may be selected from among, e.g., ammonium salts, aqueous ammonia, and gaseous ammonia, supplements as supplemented with, e.g., low levels of amino acids, vitamins, peptones, or other ingredients, The culture media may be prepared using the methods known in the art.

The transformed host cells may be tested for expression on small volume such as 5- 50 ml in LB, terrific broth or chemically defined medium. The expression may be subjected to different concentrations of inducers ranging from about 0.01 mM, about 0.05 mM, about 0.2 mM, about 0.3 mM, about 0.4 mM, about 0.5 mM, about 0.6 mM, about 0.7 mM, about 0.8 mM, about 0.9 mM and about 1 mM. The polypeptide expression is determined in an electrophoretic set up, preferably in SDS PAGE electrophoresis and viewed as over-expressed bands stained with comassie brilliant blue (-).

Subsequently, the host cells are inoculated in 500 mL flasks cultures; and allowed to grow under optimal conditions, during which CRM<sub>197</sub> expression continued from 16 hours to 32 hours. After culturing in constant agitation and preferably under aerobic conditions, the cells are harvested, and lysed. Any of known methods may be applied to

lyse cells, preferred method includes a lysis buffer containing a detergent at an appropriate concentration. After lysis, the protein component is pooled in one or more centrifugation steps. Lysis is carried out in a buffer containing Tris-HCl 20-50 mM pH 7.5-8.5, NaCl 100-150 mM, detergent 0.5-1.5% and protease inhibitor 0.5-1.5%, with agitation.

In one embodiment induction temperatures for expression is carried out between 10 to 40 °C. In one embodiment CRM<sub>197</sub> is derived at an induction temperature in a tuneable manner, wherein when induction temperature is maintained between 10 to 20 °C, more than 80% expressed CRM<sub>197</sub> is present in soluble fraction,. In another embodiment when the induction temperature is maintained between 25 to 40 °C, more than 80% of expressed CRM<sub>197</sub> obtained is in the insoluble fraction as cytoplasmic inclusion bodies, from which it is purified after a solubilisation step from the pooled cytoplasmic fraction.

In one embodiment cytoplasmic inclusion bodies are solubilized with various concentration of Urea, *per se* 1M Urea, or 2M Urea, or 3 M Urea, or 4 M Urea, or 5 M Urea, or 6 M Urea, or 7 M Urea, or 8 M Urea, or 9 M Urea.

In the specific embodiment, the yield of soluble CRM<sub>197</sub> is about 0.1 g/l, 0.25 g/L, 0.5 g/L, about 1 g/L, about 1.5 g/L, about 2 g/L, about 2.5 g/L, about 3 g/L, about 3.5 g/L, about 4 g/L, about 4.5 g/L, about 4.5 g/L, about 5 g/L.

In the specific embodiment, the yield of insoluble CRM<sub>197</sub> is about 0.25 g/L, 0.5 g/L, about 1 g/L, about 1.5 g/L, about 2 g/L, about 2.5 g/L, about 3 g/L, about 3.5 g/L, about 4 g/L, about 4.5 g/L, about 4.5 g/L, about 5 g/L.

The expressed protein is purified using ion exchange chromatographic column followed by affinity chromatography.

The invention thus involves more than one subsequent purification steps, and also exploits pI value of CRM<sub>197</sub> in an ion exchange chromatographic step, whereby it is

separated from other contaminating proteins. Finally, the quantity of CRM<sub>197</sub> is quantified by BCA/Bradford/Lowry Assay and visualised in 10-12% acrylamide gel (SDS-PAGE). The identification of polypeptide is done by Western blot and similar immunoassays. The purity and integrity of purified polypeptide is measured by SDS-PAGE and HPLC methods. The yield of the protein thus expressed is 500-1000 mg/L of the culture medium and can be subsequently varied by modulating the culture additives and conditions, as well as purification steps. The method of the invention also provides an industrially applicable method of tuning the induction time and subsequently modulating the pH and temperature of the chromatographic steps provides simple, inexpensive, and is not laborious. It excludes need of extensive steps involving preparation of buffers or kit or working solution thereof. During the removal of tag there is no need to provide additional buffers or salts or enzymes or equipments. In particular embodiment, the purified CRM<sub>197</sub> polypeptide readily lacked the first Methionine amino acid, whose presence is not desired in the final CRM<sub>197</sub> protein and removal of which entails requirement for additional purification steps. The Polypeptide thus obtained is in active and native form; it readily lacks the undesired Methionine as first amino acid without the need of additional steps. CRM<sub>197</sub> amino acid sequence was analysed by Insilico/bioinformatics tools; showed about 38.4% hydrophobicity in the protein. The isoelectric point of CRM<sub>197</sub> is found about 5.81. CRM<sub>197</sub> protein contained 4 cysteine amino acid residue and 21 proline residues. The refolding of polypeptide is confirmed by functional assays by measuring endonuclease activity over DNA. Biophysical / secondary structure confirmation is done by Circular Dichroism (CD) analysis (Fig 8) and Differential Scanning Fluorimetry (DSF) (Fig 9) of polypeptide and compared with commercially available polypeptides (Sigma Aldrich).

In another embodiment the presence of correct disulphide linkage was confirmed and compared with commercially available CRM<sub>197</sub> polypeptides (Sigma Aldrich) (Fig 10). Also the correctness of amino acid sequence of produced polypeptide was confirmed by digesting the CRM<sub>197</sub> polypeptide with multiple proteases and mapping of amino acid sequence (Fig 6). The N-terminal amino acid sequence of produced polypeptide is confirmed by Edman degradation sequencing (Fig 7).

In a preferred embodiment, the present invention provides an optimized polynucleotide sequence (SEQ ID NO. 2) and its structural variants having equal to or more than 70% similarity, preferably 85 to 99% similarity, useful for high level expression of polypeptide for CRM<sub>197</sub>.

In yet another preferred embodiment, the present invention provides an optimized polynucleotide sequence (SEQ ID NO. 2) useful for high level expression of polypeptide for CRM<sub>197</sub> in *Escherichia coli* cell.

In another preferred embodiment, the present invention provides high level expression of Diphtheria toxin or CRM<sub>197</sub> or variants thereof, using nucleic acid SEQ ID NO: 2 or a variant thereof in gram negative bacterial cell, preferably *Escherichia coli* comprising the steps of:

- a) selecting the gene SEQ ID NO: 2 or its variant thereof, which encodes polypeptide CRM<sub>197</sub>,
- b) sub cloning the gene SEQ ID NO: 2, into an expression vector,
- c) transforming the host *Escherichia coli* cell with the expression vector of step b;
- d) culturing the transformed host cell in a culture media suitable for the expression of the toxin protein;
- e) inducing the expression of fusion protein by adding IPTG as inducing agent at temperature in the range of 30 to 40 °C,
- f) extracting the bacterial toxoid in insoluble form from the host cell and
- g) purifying the CRM<sub>197</sub> in pure form with yield more than 0.5mg/l.

The purification is carried out using chromatography. The chromatography technique may be affinity chromatography, gel filtration, high pressure liquid chromatography (HPLC) or ion exchange chromatography or combination of two or more. Preferably, when CRM<sub>197</sub> is associated with tag fusion protein, affinity chromatography may be used to separate CRM<sub>197</sub> from other proteins.

In another preferred embodiment, a simple step involving a shift in temperature and pH of the column conditions also facilitate the elution of CRM<sub>197</sub> from the associated tag. More particularly, a pH in the range of 6.5-8.5 and temperature in the range of 4 °C -30 °C may be used to separate CRM<sub>197</sub> from tag.

In another embodiments, the CRM<sub>197</sub> prepared according to the present invention is used to conjugate with polysaccharide molecules isolated from *Salmonella typhi*, *Salmonella paratyphi*, *Pneumococcus*, *Haemophilus influenzae*, *Meningococcus*, *Streptococcus pneumoniae* and other pathogenetic bacteria.

In another embodiment, the CRM<sub>197</sub> prepared according to the present invention is used as a conjugated carrier for vaccines such as those against *Salmonella typhi*, *Salmonella paratyphi*, *Pneumococcus*, *Haemophilus influenzae*, *Meningococcus*, *Streptococcus pneumoniae* and other pathogenetic bacteria.

The present invention will be more specifically illustrated with reference to the following examples. However, it should be understood that the present invention is not limited by these examples in any manner but includes variations thereof within the parameters described herein, as can be known to those well-versed in the art.

### **Example 1**

#### **Step (i): Synthesis of novel CRM<sub>197</sub> gene:**

Full length CRM<sub>197</sub> gene was optimized according to *Escherichia coli* codon usage. The following parameters were used for CRM<sub>197</sub> gene optimization: Codon Usage Bias, GC content, mRNA Secondary Structure, Custom Desired Patterns, Custom Undesired Patterns, Repeat Sequences (direct repeat, inverted repeat, and dyad repeat), Restriction Enzyme Recognition Sites (deletion or insertion).

Optimized CRM<sub>197</sub> gene (SEQ ID NO. 2) was cloned at multiple cloning site of pUC57 plasmid vector using BamH1 and Sap1 restriction sites, generating pUC57\_ CRM<sub>197</sub>. The vectors containing CRM<sub>197</sub> gene was transformed in *Escherichia coli* DH5α host

and clones was selected on LB+Kanamycin<sup>r</sup> plate. The presence and correctness of CRM<sub>197</sub> gene in pUC57 was confirmed by restriction digestion of pUC57\_CRM<sub>197</sub> plasmid by Age I (located in CRM<sub>197</sub> gene) and Nde I (located in pUC57 plasmid). Further the sequence of CRM<sub>197</sub> was confirmed by PCR and DNA sequencing.

**Step (ii): Insertion of CRM<sub>197</sub> into expression vector pTWIN1**

*Escherichia coli* DH5 $\alpha$  carrying pUC57\_CRM<sub>197</sub> was grown over night in LB+Kanamycin in 50 ml volume. Bacteria was centrifuged and pellet was used for plasmid isolation. Isolation of plasmid was done by using Qiagen plasmid mini-prep kit using manufacturer instructions. Isolated plasmid was quantified by nono-drop.

CRM<sub>197</sub> (SEQ ID NO. 2) from pUC57 was excised, 5 $\mu$ g plasmid was digested with restriction endonucleases BamHI and SapI. The digested plasmid was run on 1% agarose gel and band corresponding to CRM<sub>197</sub> gene (SEQ ID NO. 2, ~1.6 kb) was purified by using Qiagen Gel extraction kit using manufacturer's instructions. Subsequently the 5  $\mu$ g of expression plasmid pTWIN1 was also digested with BamHI and SapI to generate restriction sites in it that is compatible with CRM<sub>197</sub> gene. The digested pTWIN1 was also purified from gel using Qiagen Gel extraction kit with manufacturer's instructions.

The digested CRM<sub>197</sub> gene was ligated in pTWIN1 using T4 ligase based DNA ligation kit (Promega) using manufacturer's instructions. Vector (pTWIN1) and Insert (CRM<sub>197</sub>) was mixed in 1:3, 1:4, 1:5 ratio in the presence of T4 DNA ligase and buffers in a 20 $\mu$ l reaction volume. Ligation mixture was incubated overnight at 16°C. Next morning 5 $\mu$ l of ligation mixture was added/transformed in BL21-DE3 *Escherichia coli* expression host. BL21 was transformed by using chemical transformation protocol. The ligation+ BL21 cells were incubated in ice for 30 min. After incubation heat shock was given for 45 seconds at 42°C. Sample was cooled at room temperature and 500 $\mu$ l SOC medium was added into it. The tube with transformants was incubated for 2 hours at 37°C with 200 rpm. From which 100 $\mu$ l mixture was plated on LB+Ampicillin plate for screening of transformants.

CRM<sub>197</sub> expression BL21-DE3 transformants were selected next morning from Luria Broth+Ampicillin plates. Of these 5 clones growing on Luria Broth+Ampicillin were selected and grown in 10 ml Luria Broth+Ampicillin media for overnight at 37 degrees, 200 rpm. Culture was centrifuged and plasmid was extracted from cell pellet using Qiagen plasmid extraction kit.

To verify the correctness of clone, 2µg plasmid was digested with AgeI and ApaI restriction endonuclease, respectively. AgeI site is present in CRM<sub>197</sub> and ApaI is in pTWIN1. Therefore double digestion with both the enzymes used for confirmation of correct clone. The clone was designated as pTWIN1\_ CRM<sub>197</sub> (BL21-DE3). Furthermore clones were confirmed by PCR using CRM<sub>197</sub> gene specific primers and DNA sequencing. The glycerol stock of BL21 expressing CRM<sub>197</sub> was made by growing bacteria in 10ml Luria Broth+Ampicillin overnight. Next morning 40% sterilized Glycerol was added into culture and 1ml aliquot was dispensed into cryovial. Vials were stored at -80 degree for further use in expression analysis.

**Step (iii): Confirmation of expression of CRM<sub>197</sub>:**

BL21 clone stored at -80 degrees was streaked on Luria Broth+Ampicillin plate. Plate was incubated overnight at 37 degrees. Single colony was picked up and inoculated in 50ml Luria Broth+Ampicillin media in 150ml flask. Flask was incubated at 37 degrees, 200 RPM until OD<sub>600</sub>= 1. Once OD reaches to desired point, 5 ml culture was drawn which is used as uninduced culture. Uninduced culture was kept on ice until use. To induce the expression of CRM<sub>197</sub> gene 0.5mM IPTG was added to remaining 45 ml culture and flask was further incubated for additional 4 hours at 30 degree and 200 rpm rotation. Induced culture was harvested after 4 hours and expression of CRM<sub>197</sub> was examined by SDS-PAGE (Figure 1) and Western Blot (Figure 2).

For SDS-PAGE analysis 1ml culture of induced and uninduced culture (both normalized for OD<sub>600</sub>=1) was taken into 1.5ml Eppendorf tube. The tube was centrifuged and pellet was resuspended into 50µl PBS. In this suspension 50 µl SDS-PAGE loading buffer with reducing agent (2x) was added. The mixture was boiled at 100 degree for 5 min.

Sample was cooled at room temperature and 20 µl of uninduced and induced culture was loaded in the 4-12% Tris Glycine gel. The gel was run for 1.5 hours at 150 volts. Gels were taken out and incubated in Coomassie Brilliant Blue dye for 1 hour. After staining gel was destained in destaining solution containing 40% methanol=10 % acetic acid for 3 hours. The CRM<sub>197</sub> expression was visualized as ~58KD protein that is only visible in induced culture. For western blot a separate set of gel was run in the same manner as SDS-PAGE and gel was blotted on PVDF (polyvinylidene difluoride membrane). The membrane was immunoblotted by anti- CRM<sub>197</sub> antibody. In the western blot CRM<sub>197</sub> appeared as single immunoreactive band at ~58 Kd. No CRM<sub>197</sub> specific band was observed in uninduced culture. This experiment confirms that the clone generated in the present study can express rCRM<sub>197</sub> protein. These clone were further used for large scale production and purification of CRM<sub>197</sub>.

**Step (iv): Fermentation and Purification of CRM<sub>197</sub> from BL21 *Escherichia coli***

One ml vial of BL21 *Escherichia coli* cells was inoculated into 50 ml LB+Amp media and grown overnight at 37 degrees, 200 rpm. Fermentation was done at 20L scale. *Escherichia coli* cells were inoculated to the fermenter and cultivated at 30 degrees centigrade. The culture was induced with 0.5 mM IPTG at OD<sub>600</sub>=20. After 12 hours post induction fermentation culture was harvested and cell pellet was prepared by centrifugation. Cell pellet was lysed mechanically in homogenizer. Inclusion body (which contains the desired protein CRM<sub>197</sub>) was isolated by centrifugation of cell lysates. Supernatant was discarded and pellet was retained which contains Inclusion body (IBs). IBs were homogenized by resuspending pellet in 8M urea and protein was purified by ion exchange chromatography. Quantification of CRM<sub>197</sub> at fermentation level was measured. Whole cells lysates was run on SDS-PAGE along with the known amount of BSA as standard. The quantification of CRM<sub>197</sub> which appeared as ~58 KD band in SDS-PAGE (Figure 3) was quantified by densitometry analysis. BSA was used as reference for quantification. The total amount of protein was found as 1.4 gram CRM<sub>197</sub>/litre of BL21 *Escherichia coli* cells.

Solubilization test of the CRM<sub>197</sub> prepared above was carried out in Electrophoretic gel run (SDS PAGE 12%) showing test conducted on solubilization of CRM<sub>197</sub>, Lane 1: Urea solubilized fraction of CRM<sub>197</sub>; Lane 2: Supernatant; Lane 3: Sample from first pellet wash; Lane 4: Sample pooled from 2<sup>nd</sup> pellet wash (Figure 4).

Presence of CRM<sub>197</sub> in the sample was determined by Size exclusion chromatography (SEC-HPLC) wherein major eluted peak shows the presence of CRM<sub>197</sub> in the sample (Figure 5).

Primary amino acid sequence of the CRM<sub>197</sub> prepared above was determined by Peptide mass fingerprint (mass spectrometry) and had 100% sequence similarity with the reference CRM<sub>197</sub> sequence (Figure 6).

- A. Peptic digest of BE rCRM analysed by LC-MS.
- B. Sequence coverage in Trypsin, Glu-C and Asp-N digest CRM<sub>197</sub> (identical to SEQ ID NO. 1 i.e. shows 100% homology)

N-Terminal sequence of rCRM<sub>197</sub> prepared above was confirmed by Edman degradation. The 10 amino acid sequence GADDVVDSSK (N-Term acetyl) shows starting portion of purified polypeptide. The first amino acid is identified as G (Figure 7).

CD spectra of the CRM<sub>197</sub> prepared above was overlapped with the reference CRM<sub>197</sub>. The result shows that recombinant CRM<sub>197</sub> prepared above is structurally similar to the reference CRM<sub>197</sub> (Figure 9).

Confirmation of disulphide bonds of CRM<sub>197</sub> prepared by the above method was analysed (Figure 10) and was confirmed by mass spectrometry that two disulphide bonds present in the protein first links amino acid 186 to 201 and second bond links amino acid 461 to 471.

**Table I**

Cystine Numbering	Peptide sequence
186	GQDAMYEYMAQACAGNR (SEQ ID NO. 11)
201	SVGSSLSCINLDWDVIR (SEQ ID NO. 12)
461	CR (SEQ ID NO. 13)
471	AIDGDVTFCRPK (SEQ ID NO. 14)

**Table II**

Tryptic peptides chain combination	Theoretical Mono m/z	Observed Mono m/z
Cys 186-201	935.6706	935.6694
Cys 461-471	913.9558	913.9539

Antigenic similarity of CRM<sub>197</sub> prepared by the above method with reference CRM<sub>197</sub> was confirmed by CRM<sub>197</sub> specific ELISA. All the CRMs coming from difference source showed similar recognition profile with monoclonal antibodies (Figure 11).

**We claim:**

1. An optimized polynucleotide sequence comprising of SEQ ID NO. 2 and its variants which are at least 70% homologous to the said optimized polynucleotide sequence SEQ ID NO. 2.
2. An optimized polynucleotide sequence comprising of SEQ ID NO. 2 and its variants which are at least 70 to 88 % homologous to the said optimized polynucleotide sequence SEQ ID NO. 2
3. Variants of SEQ ID NO. 2 as claimed in claims 1 and 2, selected from SEQ ID NO. 3, 4, 5, 6, 7, 8, 9, 10.
4. A process for the production of polypeptide, comprising steps of:
  - a) selecting an optimized polynucleotide sequence essentially consisting of SEQ ID NO. 2 or its variants which are at least 70% homologous to SEQ ID NO.2,
  - b) optionally ligating the polynucleotide sequences of step (a) into a suitable vector,
  - c) inserting or transforming the polynucleotide sequence into *Escherichia coli* host cell,
  - d) culturing the transformed host cell in a culture media for high level expression of the polypeptide,
  - e) maintaining the induction temperature between 10 to 40 °C to produce polypeptide,
  - f) extracting the bacterial polypeptide from the host cell, followed by purification to obtain pure polypeptide with high yields.
5. The process as claimed in claim 4, wherein suitable vector is a plasmid vector selected from pET9a, pET3a, pET3b, pET3c, pET5a, pET5b, pET5c, pET9b, pET9c, pET12a, pTWIN1, pTWIN2, pET12b, pET12c, pET17b.

6. The process as claimed in claim 4, wherein *Escherichia coli* strain is selected from BL21 (DE3), BL21 A1, HMS174 (DE3), DH5 $\alpha$ , W3110, B834, origami, Rosetta, NovaBlue (DE3), Lemo21 (DE3), T7, ER2566 and C43 (DE3).
7. The process as claimed in claim 4, wherein the yield of CRM<sub>197</sub> is about 0.1 g/l, 0.25 g/L, 0.5 g/L, about 1 g/L, about 1.5 g/L, about 2 g/L, about 2.5 g/L, about 3 g/L, about 3.5 g/L, about 4 g/L, about 4.5 g/L, about 4.5 g/L, about 5 g/L.
8. The process as claimed in claim 4, wherein periplasmic expression is brought about by providing a suitable induction temperature for the expression of polynucleotide, without any heterologous sequence for directed transport into the periplasmic space.
9. The process as claimed in claim 8, wherein the induction temperature ranges from 10 to 40 °C.
10. The process as claimed in claim 4, wherein the polypeptide is a carrier protein.
11. The carrier protein as claimed in claim 10, wherein the carrier protein is Diphtheria toxin mutant CRM<sub>197</sub>.
12. CRM<sub>197</sub> as claimed in any of the preceding claims, wherein the said CRM<sub>197</sub> is conjugated with polysaccharide molecules isolated from *Salmonella typhi*, *Salmonella paratyphi*, *Pneumococcus*, *Haemophilus influenzae*, *Meningococcus*, *Streptococcus pneumoniae* and other pathogenetic bacteria.
13. CRM<sub>197</sub> as claimed in any of the preceding claims, wherein the said CRM<sub>197</sub> is used as a conjugated carrier for vaccines such as those against *Salmonella typhi*, *Salmonella paratyphi*, *Pneumococcus*, *Haemophilus influenzae*, *Meningococcus*, *Streptococcus pneumoniae* and other pathogenetic bacteria.

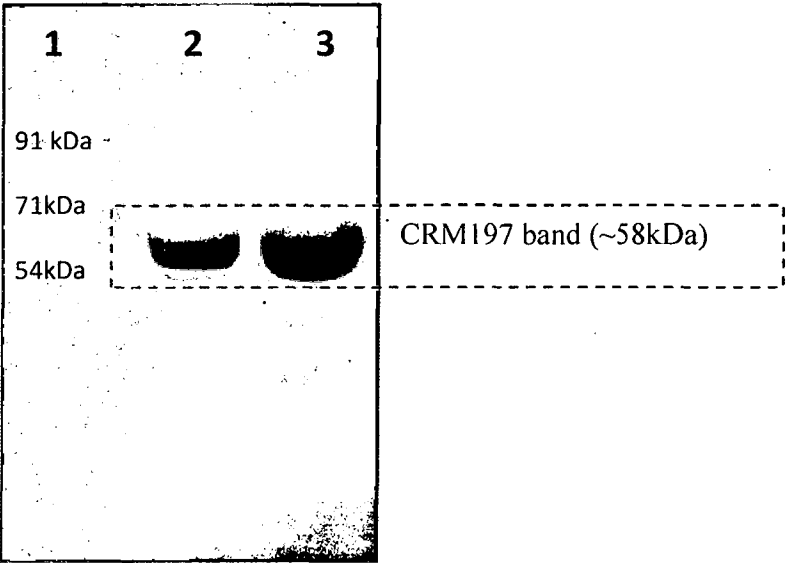


FIGURE 1

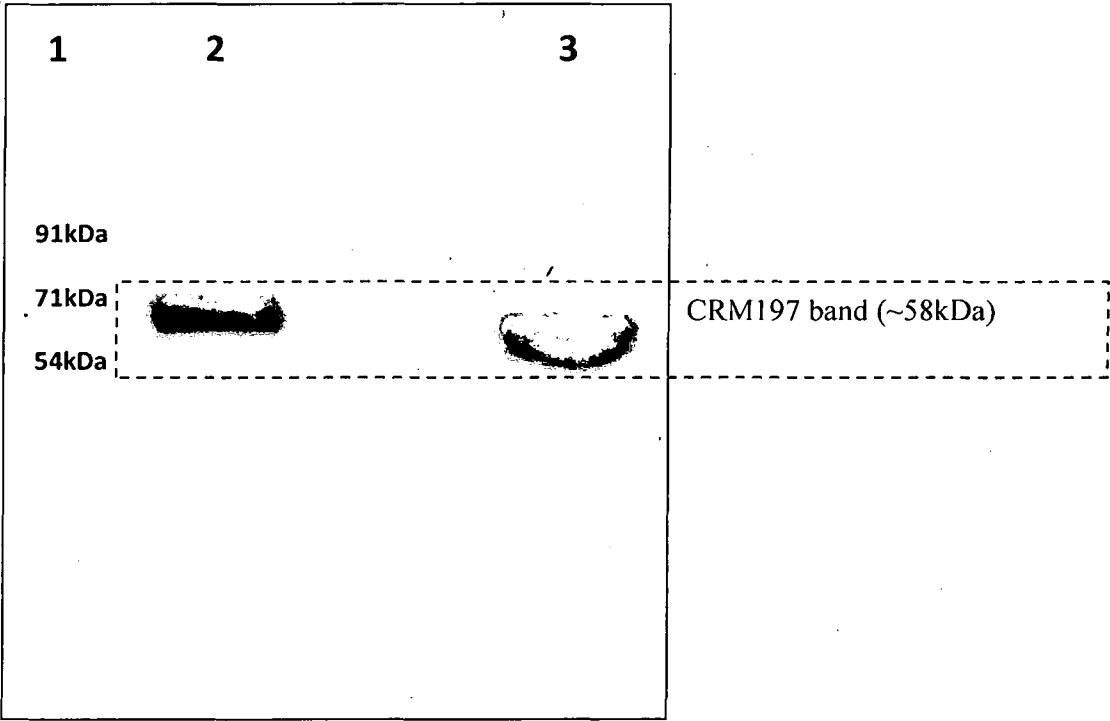


FIGURE 2

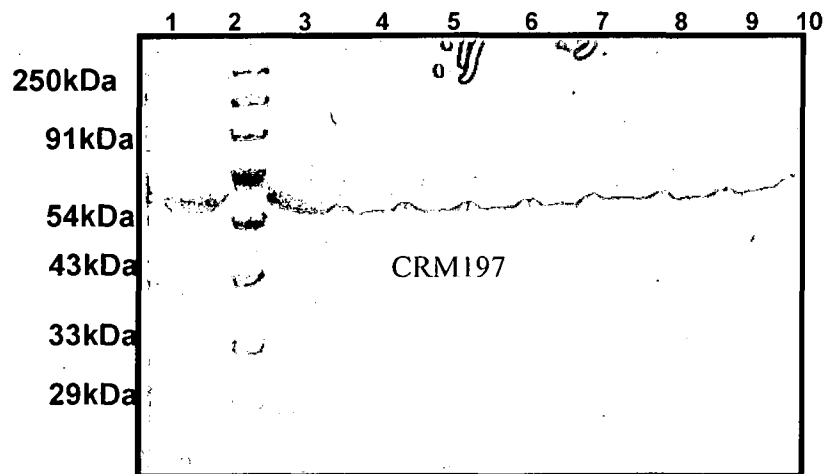


FIGURE 3

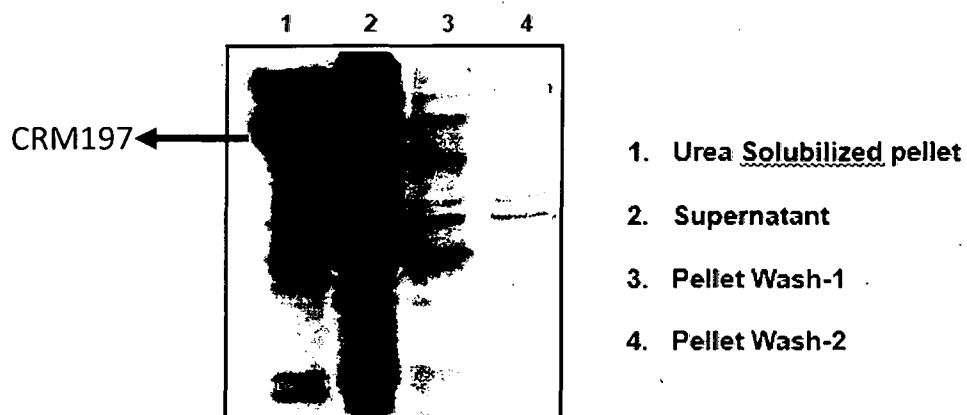


FIGURE 4

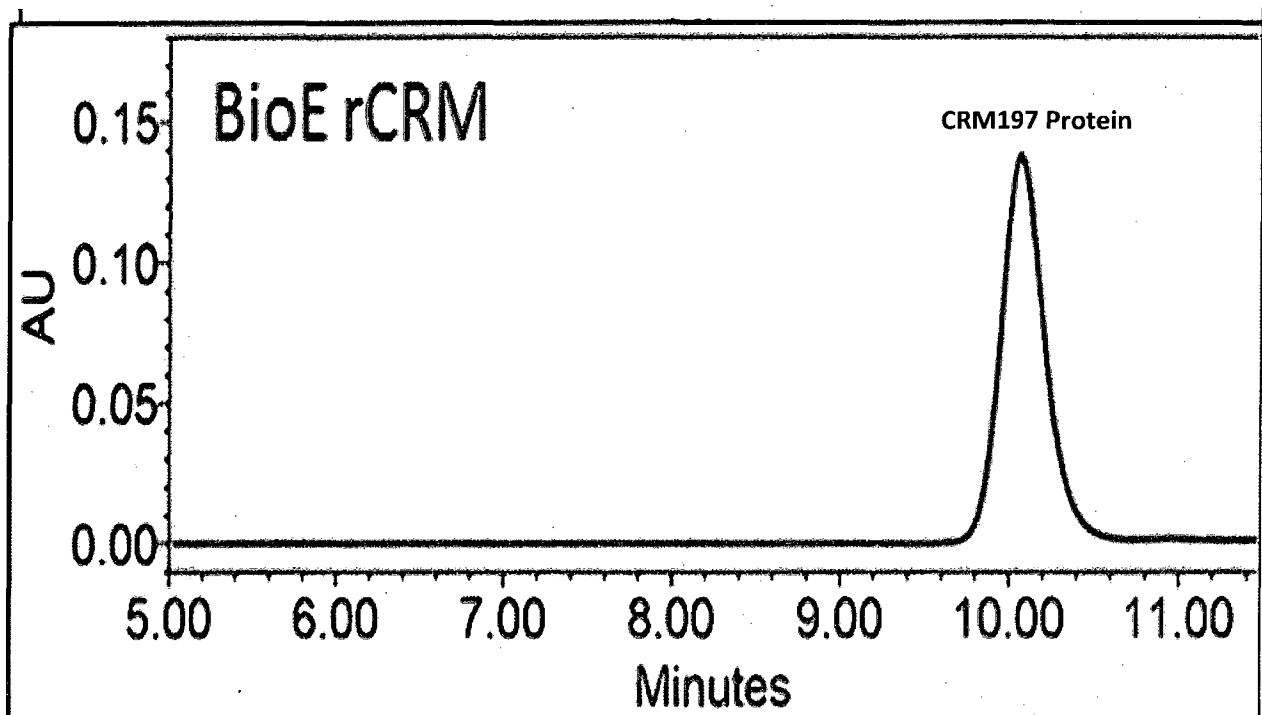


FIGURE 5

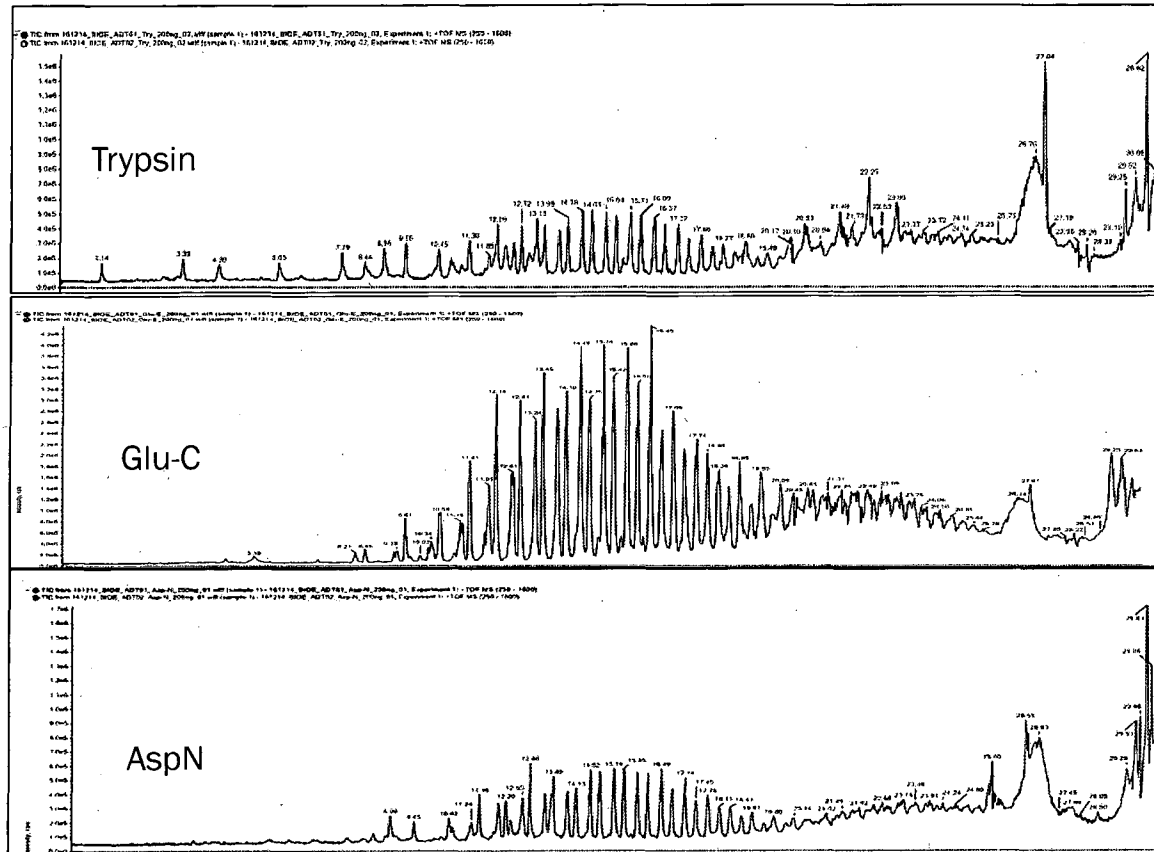


FIGURE 6

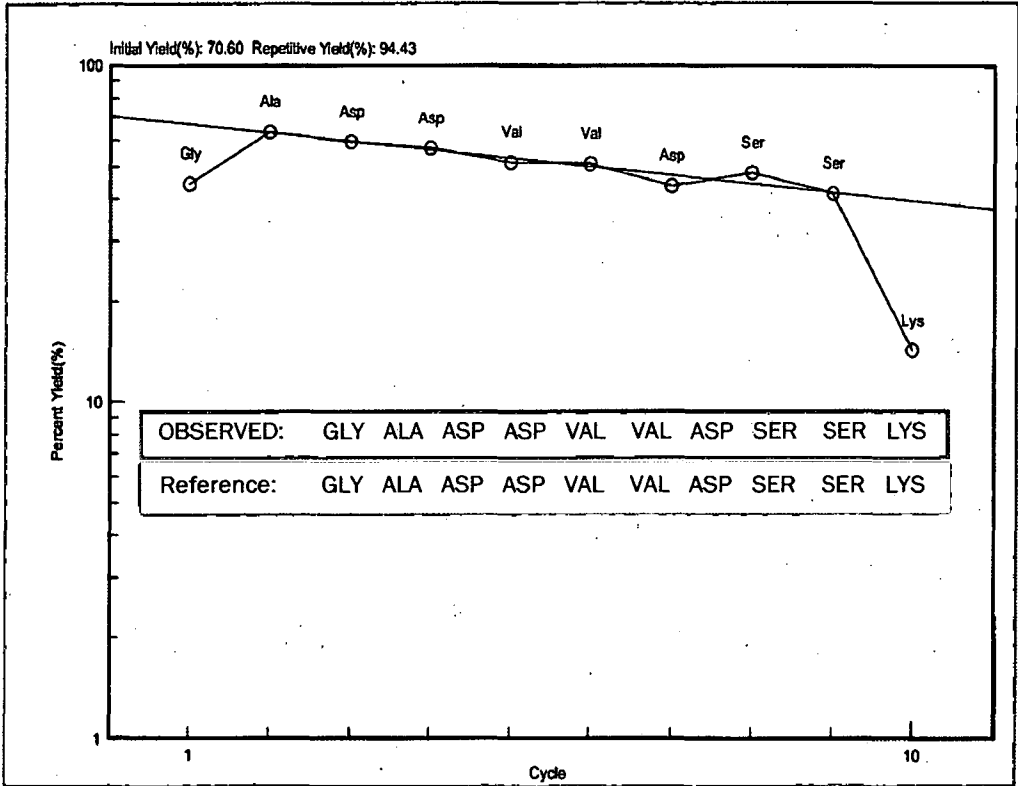


FIGURE 7

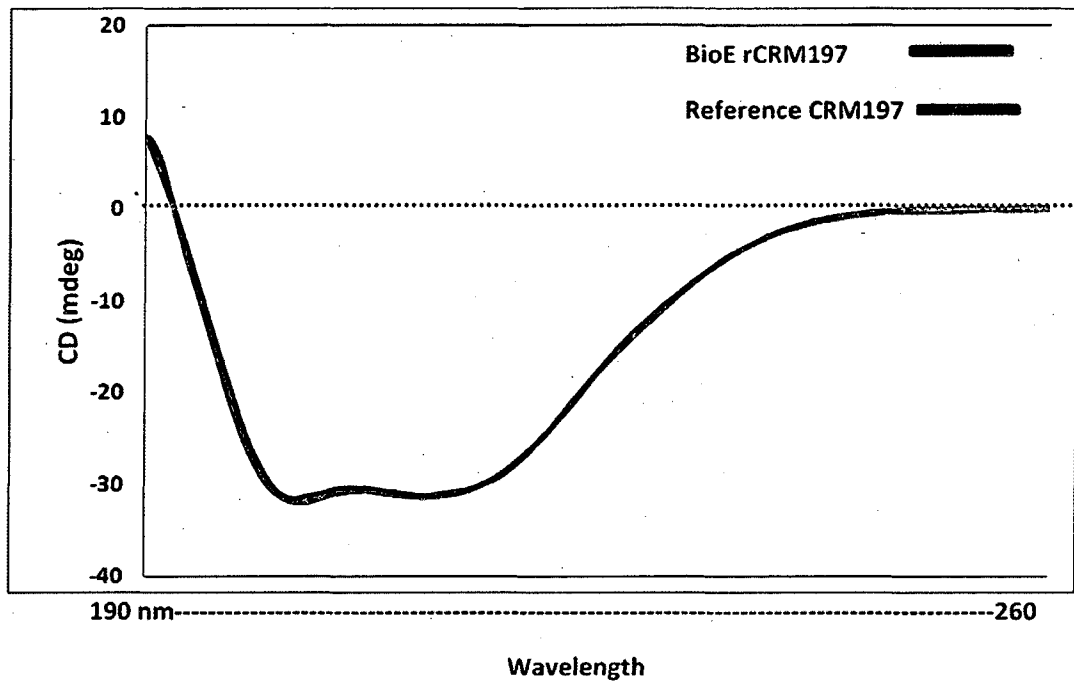


FIGURE 8

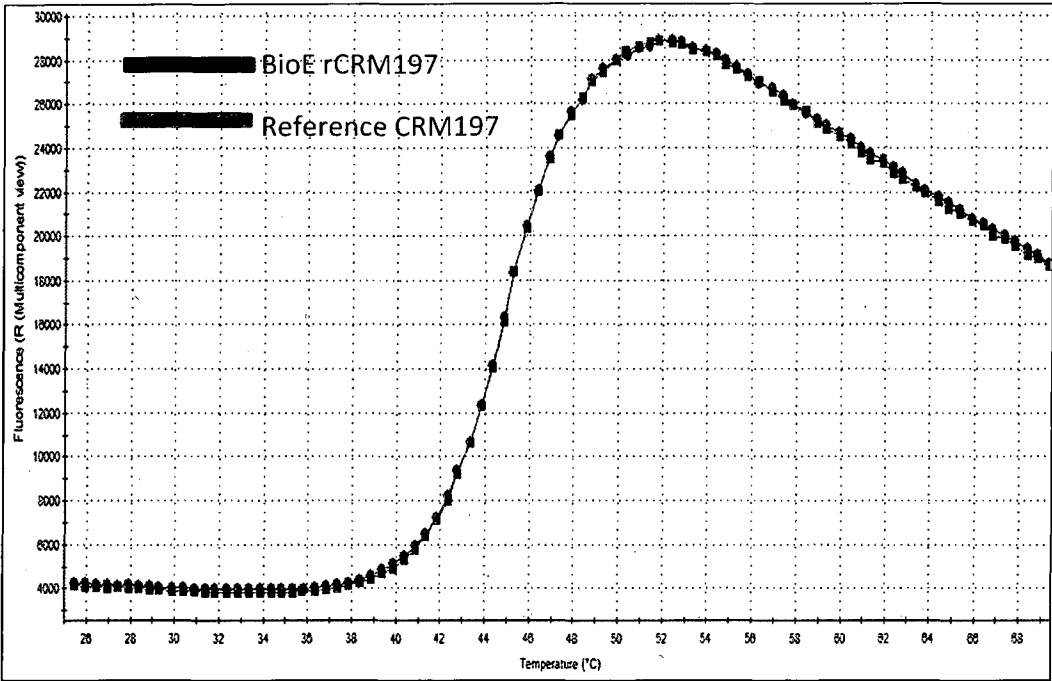


FIGURE 9

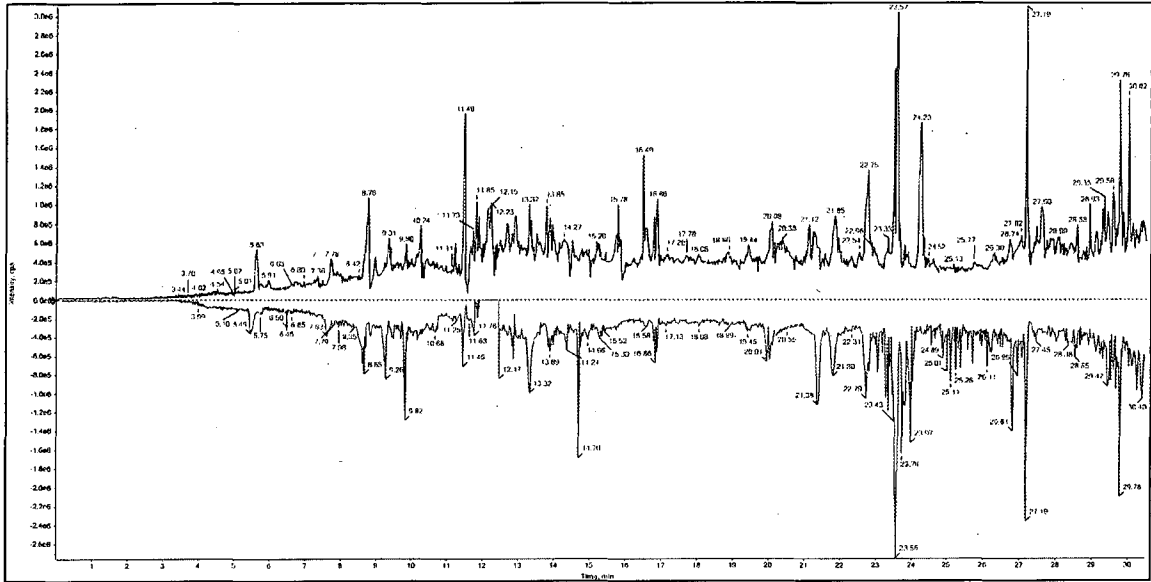


FIGURE 10

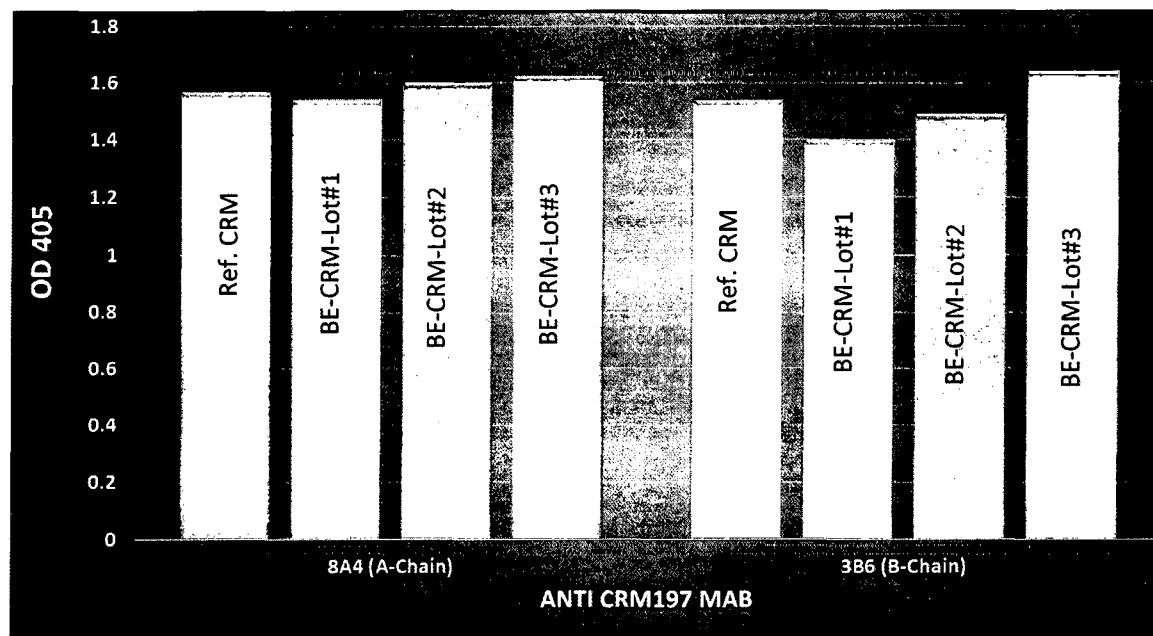


FIGURE 11

## INTERNATIONAL SEARCH REPORT

International application No

PCT/IN2015/000427

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. C07K14/34 C12N15/10 C12N15/70 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) C07K C12N  Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, BIOSIS, EMBASE, WPI Data		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103 266 125 A (BEIJING MINHAI BIOTECHNOLOGY CO LTD) 28 August 2013 (2013-08-28) abstract paragraphs [0003], [0019], [0028], [0029] claim 9 sequence 1  -----  -/--	1-13
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p>		
Date of the actual completion of the international search  20 April 2016		Date of mailing of the international search report  03/05/2016
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer  Chavanne, Franz

## INTERNATIONAL SEARCH REPORT

International application No

PCT/IN2015/000427

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>ALESSANDRA STEFAN ET AL: "Overexpression and purification of the recombinant diphtheria toxin variant CRM197 in", JOURNAL OF BIOTECHNOLOGY, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 156, no. 4, 15 August 2011 (2011-08-15), pages 245-252, XP028119257, ISSN: 0168-1656, DOI: 10.1016/J.JBIOTEC.2011.08.024 [retrieved on 2011-08-25] abstract page 246 - page 248 -----</p>	1-11,13

## INTERNATIONAL SEARCH REPORT

### Information on patent family members

International application No

PCT/IN2015/000427

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CN 103266125	A	28-08-2013	NONE
-----			