



(86) Date de dépôt PCT/PCT Filing Date: 2015/02/02

(87) Date publication PCT/PCT Publication Date: 2015/08/06

(45) Date de délivrance/Issue Date: 2021/01/12

(85) Entrée phase nationale/National Entry: 2016/07/28

(86) N° demande PCT/PCT Application No.: US 2015/014130

(87) N° publication PCT/PCT Publication No.: 2015/117093

(30) Priorité/Priority: 2014/01/31 (US61/934,377)

(51) Cl.Int./Int.Cl. *C12N 15/64* (2006.01),
C07K 1/16 (2006.01), *C07K 14/34* (2006.01),
C12N 1/21 (2006.01), *C12N 15/31* (2006.01),
C12N 15/53 (2006.01), *C12N 15/63* (2006.01),
C12N 5/00 (2006.01), *C12N 5/02* (2006.01),
C12P 21/02 (2006.01), *G01N 33/53* (2006.01)

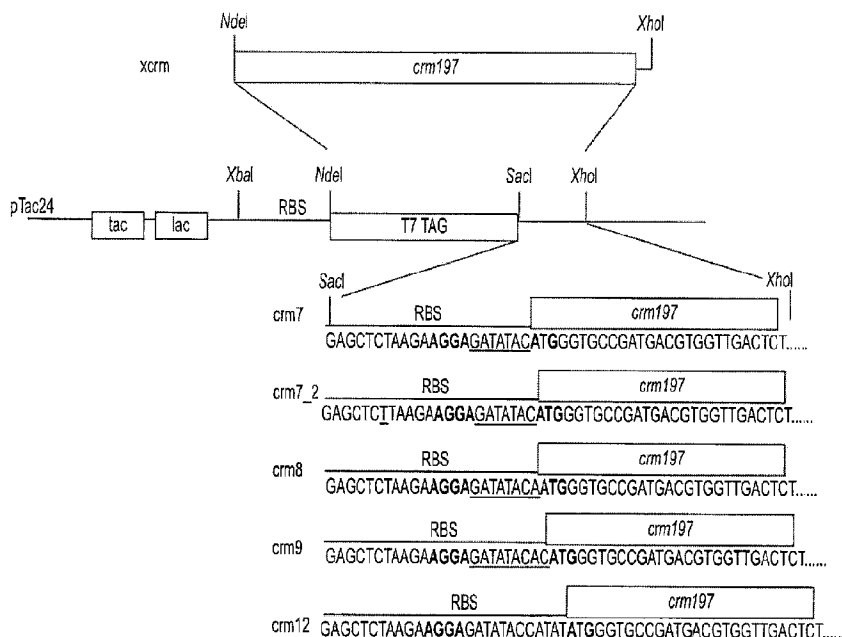
(72) Inventeurs/Inventors:
OGANESYAN, NATALIA, US;
LEES, ANDREW, US

(73) Propriétaire/Owner:
FINA BIOSOLUTIONS, LLC, US

(74) Agent: MBM INTELLECTUAL PROPERTY LAW LLP

(54) Titre : EXPRESSION ET PURIFICATION DE CRM197 ET DE PROTEINES ASSOCIEES

(54) Title: EXPRESSION AND PURIFICATION OF CRM197 AND RELATED PROTEINS



(57) Abrégé/Abstract:

The present invention is directed to the cells, compositions and methods for the production of recombinant protein. In particular, the invention is directed to a production process for obtaining high levels of soluble recombinant CRM₁₉₇ protein from E. coli. Cells preferably contain one or more mutations of disulfide reductase genes, so that disulfide reductase activity is reduced. The invention also relates to purification method for CRM₁₉₇ as well as characterization of properly folded CRM₁₉₇ protein.

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

(19) World Intellectual Property
Organization
International Bureau(10) International Publication Number
WO 2015/117093 A1(43) International Publication Date
6 August 2015 (06.08.2015)

(51) International Patent Classification:

C12N 15/64 (2006.01) C12N 5/02 (2006.01)
C12N 15/00 (2006.01) C12P 21/06 (2006.01)
C12N 5/00 (2006.01)

(21) International Application Number:

PCT/US2015/014130

(22) International Filing Date:

2 February 2015 (02.02.2015)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

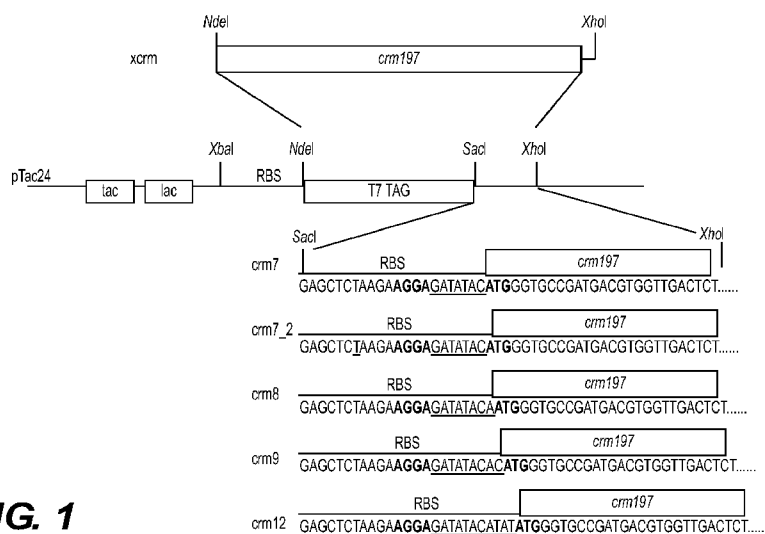
61/934,377 31 January 2014 (31.01.2014) US

(71) Applicant: FINA BIOSOLUTIONS, LLC [US/US];
9430 Key West Ave., Suite 200, Rockville, Maryland
20850 (US).(72) Inventors: OGANESYAN, Natalia; 11413 Flints Grove
Lane, North Potomac, Maryland 20878 (US). LEES, An-
drew; 1910 Glen Ross Rd, Silver Spring, Maryland 20910
(US).(74) Agents: REMENICK, James et al.; Remenick PLLC,
1025 Thomas Jefferson St., NW, Suite 175, Washington,
District of Columbia 20007 (US).(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).

Published:

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

(54) Title: EXPRESSION AND PURIFICATION OF CRM197 AND RELATED PROTEINS

**FIG. 1**(57) Abstract: The present invention is directed to the cells, compositions and methods for the production of recombinant protein. In particular, the invention is directed to a production process for obtaining high levels of soluble recombinant CRM₁₉₇ protein from *E. coli*. Cells preferably contain one or more mutations of disulfide reductase genes, so that disulfide reductase activity is reduced. The invention also relates to purification method for CRM₁₉₇ as well as characterization of properly folded CRM₁₉₇ protein.

EXPRESSION AND PURIFICATION OF CRM197 AND RELATED PROTEINS

5 Background

1. Field of the Invention

The present invention relates to the field of recombinant protein production in bacterial hosts. In particular, the present invention relates to a production process for obtaining high levels of soluble recombinant CRM₁₉₇ protein from *E. coli*. The invention also relates to
10 purification and characterization methods for CRM₁₉₇ as well as uses of the CRM₁₉₇ produced by the method.

2. Description of the Background

Diphtheria toxin (DT) is a proteinaceous exotoxin synthesized and secreted by pathogenic strains of *Corynebacterium diphtheriae*. These pathogenic strains contain a
15 bacteriophage lysogen that carries the toxin gene. Diphtheria toxin is an ADP-ribosylating enzyme that is secreted as a proenzyme of 535 residues and processed by trypsin-like proteases with release of two fragments (A and B). Fragment A uses NAD as a substrate, catalyzing the cleavage of the N-glycosidic bond between the nicotinamide ring and the N-ribose and mediating the covalent transfer of the ADP-ribose (ADPRT activity) to the modified histidine 715
20 (diphthamide) of the elongation factor EF-2. This post-translational diphthamide modification inactivates EF-2, halting protein synthesis and resulting in cell death. The A fragment of DT (also named C domain) carries the catalytic active site and is the only fragment of the toxin required for the final step of intoxication. The R domain, carried on the B fragment, mediates binding to receptors on the host cell surface and the T domain, also carried on the B fragment,
25 promotes the pH-dependent transfer of fragment A to the cytoplasm. An Arginine-rich disulfide-linked loop connects fragment A to fragment B (or domain C to domains TR). This inter-chain disulfide bond is the only covalent link between the two fragments after proteolytic cleavage of the chain at position 186. The isolation of various non-toxic and partially toxic immunologically cross-reacting forms of diphtheria toxins (CRMs or cross reacting materials) resulted in
30 discovery of CRM₁₉₇ (Uchida et al., Journal of Biological Chemistry 248, 3845-3850, 1973; see

also Giannini et al. Nucleic Acids Res. 1984 May 25;12(10):4063-9). Preferably, CRMs can be of any size and composition that contain all or a portion of DT.

CRM₁₉₇ is a largely enzymatically inactive and nontoxic form of diphtheria toxin that contains a single amino acid substitution G52E. This mutation causes intrinsic flexibility of the active-site loop in front of the NAD-binding site and reduces the ability of CRM₁₉₇ to bind NAD and eliminates toxic properties of DT (Malito et al., Proc Natl Acad Sci USA 109(14):5229-342012) Like DT, CRM₁₉₇ has two disulfide bonds. One disulfide joins Cys186 to Cys201, linking fragment A to fragment B. A second disulfide bridge joins Cys461 to Cys471 within fragment B. Both DT and CRM₁₉₇ have fragment A-associated nuclease activity (Bruce et al., Proc. Natl. Acad. Sci. USA 87, 2995-8, 1990).

Many antigens are poorly immunogenic, especially in infants, unless chemically linked to a protein ("conjugation"), thereby forming a conjugate or conjugate vaccine. The protein component of these conjugate vaccines is also called the "carrier protein". CRM₁₉₇ is commonly used as the carrier protein for protein-carbohydrate and hapten-protein conjugates. As a carrier protein, CRM₁₉₇ has a number of advantages over diphtheria toxoid as well as other toxoid proteins, many of which have been documented (Shinefield Vaccine, 28:4335, 2010, Broker et al, Biologicals, 39:195 2011). For example since CRM₁₉₇ is genetically detoxified, it retains a larger complement of lysines, which are used for conjugation but are blocked by chemical toxoiding. CRM₁₉₇ has proven to be an effective carrier protein for *Streptococcus pneumoniae* capsular polysaccharides, as evidenced by the success of PREVNAR™ (Pfizer), a vaccine consisting of up to 13 capsular polysaccharides chemically linked to CRM₁₉₇. There is also evidence suggesting that compared with tetanus toxoid, there is less carrier-induced suppression of the immune response, especially when there are many individual polysaccharides linked to the same carrier protein.

CRM₁₉₇ and native DT have a similar affinity for the diphtheria toxin receptor (DTR), which has an identical amino acid sequence to the HB-EGF precursor pro-HB-EGF (Mitamura et al., J. Biol. Chem. 272(43):27084-90, 1997). CRM₁₉₇ binds to the soluble form of HB-EGF, as well as to the membrane form pro-HB-EGF, and inhibits HB-EGF mitotic action by preventing its binding to EGF receptor. Thus CRM₁₉₇ may also have a future role in cancer therapy (Miyamoto et al., Anticancer Res. Nov-Dec 27(6A):3713-21, 2007).

CRM₁₉₇ has been produced in the original host *Corynebacterium*, but yields are low, typically <50mg/L and, in addition, *Corynebacterium* growth is relatively slow as compared with, for example, *E. coli*. There are proprietary strains of *Corynebacterium* that have been engineered to produce CRM₁₉₇ at higher levels (U.S. Patent No. 5,614,382). CRM₁₉₇ has also
5 been expressed in a proprietary strain of *Pseudomonas fluorescens* and expressed at high levels. Production of CRM₁₉₇ in *E. coli* would be advantageous since *E. coli* is a BL1 level organism that is inexpensive to culture and propagate. Production of CRM₁₉₇ in *E. coli* has mainly resulted in insoluble inclusion bodies (generally insoluble), which then requires a difficult refolding process, resulting in low yields (EP20100742260) or with an additional peptide sequence (a tag) (J
10 Biotechnol. 2010 Dec 20;156(4):245-52, Overexpression and purification of the recombinant diphtheria toxin variant CRM197 in Escherichia coli. Stefan A, Conti M, Rubboli D, Ravagli L, Presta E, Hochkoeppler A. A method for the overexpression of soluble tag free CRM₁₉₇ in *E. coli* suitable for the large quantity protein production, has not been reported. Thus, there is a need for better methods to produce CRM₁₉₇ in an efficient and cost-effective manner.

15 Summary of the Invention

The present invention overcomes the problems and disadvantages associated with current strategies and designs and provide new compositions and methods for producing CRM.

One embodiment of the invention is directed to methods of producing all or a portion of a CRM protein comprising: providing a recombinant cell that contains an expression vector that
20 contains an inducible promoter functionally linked to a polycistronic genetic sequence wherein at least one cistron encodes the CRM protein; inducing the expression vector to produce CRM protein; and isolating the CRM protein expressed. Preferably the recombinant cell has a reduced activity of one or more disulfide reductase enzymes and also preferably, each cistron contains a ribosome binding site and an initiation codon. Preferably the polycistronic genetic sequence
25 contains at least one spacer between one or more ribosome binding sites and one or more initiation codons. Preferably the CRM protein expressed by the cell is soluble and also preferably, the CRM protein expressed is intracellular, periplasmic or secreted. Preferably the recombinant cell is propagated at a temperature from about 15°C to about 32°C and also preferably, the CRM protein is isolated from the cell by chromatography. Preferable
30 chromatography media include, for example, a dextran sulfate resin, a gel resin, an active

sulfated resin, a phosphate resin, a heparin resin or a heparin-like resin. Another embodiment of the invention comprises CRM protein isolated by the methods of the invention.

Another embodiment of the invention is directed to methods of producing all or a portion of a CRM protein, such as preferably CRM₁₉₇, comprising; providing a recombinant cell that
5 contains an expression vector, wherein the recombinant cell has been modified to shift the redox status of the cytoplasm to a more oxidative state as compared to an unmodified recombinant cell and the expression vector contains an inducible promoter functionally linked to a CRM coding sequence, a spacer sequence between a ribosome binding site and an ATG codon, an expression enhancer region upstream of the CRM coding sequence; inducing the expression vector to
10 produce CRM protein; and isolating the CRM protein expressed. The recombinant cell may be a eukaryotic cell or a prokaryotic cell. Preferably the recombinant cell is a prokaryotic cells such as, for example, an *E. coli* cell or a derivative or strain of *E. coli*. Preferably, the recombinant cell modification comprises a reduced activity of one or more disulfide reductase enzymes such as, for example, one or more of an oxidoreductase, a dihydrofolate reductase, a thioredoxin and
15 thioredoxin reductase, a protein reductase or a glutathione reductase. Preferably the reduced activity of the one or more disulfide reductase enzymes shifts the redox state of the cytoplasm of the recombinant cell to an oxidative state as compared with a non-recombinant cell. Preferably the CRM coding sequence encodes one or more CRM epitopes, CRM peptide sequences, CRM domains, or combinations thereof. Preferably the spacer comprises more or less than 9
20 nucleotides such as, for example, between 5 and 20 nucleotides. Preferably the expression enhancer comprises a ribosome binding site upstream of the CRM coding sequence and an ATG codon. Preferably the CRM protein expressed by the cell is soluble and is intracellular, periplasmic or secreted. Preferably the recombinant cell is propagated at a temperature from about 15°C to about 32°C.

25 Preferably, the CRM protein is isolated from the cell by chromatography comprising, as a preferable chromatography medium, a dextran sulfate resin, an active sulfate resin, a phosphate resin, a heparin resin or a heparin-like resin.

Another embodiment of the invention is directed to CRM protein isolated by the methods of the invention. Preferably, the isolated CRM protein is conjugated and the conjugated CRM
30 protein is formulated as a vaccine.

Another embodiment of the invention is directed to methods of producing all or a portion of a CRM protein such as for example a protein or peptide produced from a CRM coding sequence that encodes one or more CRM epitopes, CRM peptide sequences, CRM domains, or combinations thereof, and preferably CRM₁₉₇, comprising providing a recombinant cell that
5 contains an expression vector, wherein the expression vector contains a promoter functionally linked to an EES coding sequence preceded by a ribosome binding site; expressing CRM protein from the CRM coding sequence preceded by a ribosome binding site; and isolating the CRM protein expressed. Preferably the recombinant cell is a prokaryotic or a eukaryotic cell and preferably the prokaryotic cell is an *E. coli* cell or a derivative or strain of *E. coli*. Preferably the
10 promoter is constitutive or inducible. Preferably the recombinant cell has been modified to shift the redox status of the cytoplasm to a more oxidative state as compared to an unmodified recombinant cell. Preferably the modified recombinant cell has reduced activity of one or more of an thiol-disulfide oxidoreductases, and or enzymes involved in the thioredoxin and glutaredoxin systems (e.g. thioredoxin, thioredoxin reductase, glutathione reductase) Preferably
15 the expression vector contains controlled by ribosome binding site an expression enhancer sequence (e.g., SEQ ID 15) or such as, for example including T7 tag sequence, upstream ribosome binding site upstream of the CRM coding sequence.

Another embodiment of the invention is directed to methods for isolating and/or purifying CRM protein comprising: loading the CRM protein onto a chromatography column
20 containing a resin with a loading buffer wherein the resin is preferably a dextran sulfate resin, a, an active sulfate resin, a phosphate resin, a heparin resin or a heparin-like resin; washing the resin with one or more washing buffers; and eluting CRM protein from the resin with an elution buffer. Preferably the loading buffer and the washing buffer are or contain the same components and at the same or in similar amounts. Preferably the loading buffer and the one or more
25 washing buffers are low conductivity buffers such as, for example, Tris-HCl, HEPES, sodium phosphate buffers a conductivity of about 10 mS/cm or less (e.g., 1 mS/cm, 2 mS/cm, 3 mS/cm, 4 mS/cm, 5 mS/cm, 6 mS/cm, 7 mS/cm, 8 mS/cm, 9 mS/cm). Preferably the elution buffer is a high conductivity buffer such as, for example, buffers with added salts such for example, NaCl, or KCl, at a conductivity of about 10 mS/cm or more (e.g., 12 mS/cm, 14 mS/cm, 15 mS/cm, 20
30 mS/cm, 25 mS/cm, 30 mS/cm, 40 mS/cm, 50 mS/cm, 60 mS/cm, 70 mS/cm, 80 mS/cm, 90 mS/cm, 100 mS/cm or more).

Another embodiment of the invention is directed to methods of characterizing folding of diphtheria toxin or CRM protein comprising: contacting diphtheria toxin or CRM protein to HB-EGF; determining the amount of binding of diphtheria toxin or CRM protein to HB-EGF; and determining the folding of diphtheria toxin or CRM protein by the amount of binding
 5 determined, wherein binding indicates correct folding. Preferably the diphtheria toxin or CRM contains a receptor binding domain. Preferably the CRM protein comprises CRM₁₉₇. Also preferably the at least one of the diphtheria toxin or CRM protein and/or the HB-EGF is bound to a solid support. Preferably the amount of binding of diphtheria toxin or CRM protein to HB-EGF is determined by an ELISA and the CRM protein that binds to HB-EGF is soluble in PBS.

10 Another embodiment of the invention comprises expression vectors that comprise a promoter and two or more cistrons at least one encoding a protein, wherein at least one cistron encodes CRM protein and each cistron has a ribosome binding site and an initiation codon. Preferably the expression vector further comprising a spacer between the ribosome binding site and the initiation codon and also preferably the spacer comprises from 5 to 20 nucleotides.

15 Preferably the spacer does not comprise 9 nucleotides.

Other embodiments and advantages of the invention are set forth in part in the description, which follows, and in part, may be obvious from this description, or may be learned from the practice of the invention.

Description of the Drawings

20 Figure 1 Schematic of vector constructs crm7 (SEQ ID NO 10), crm7_2 (SEQ ID NO 11), crm 8 (SEQ ID NO 12), crm 9 (SEQ ID NO 13), and crm 12 (SEQ ID NO 14) with the ribosome binding site (RBS) and start codon (ATG) both indicated in bold with the spacer sequence underlined.

Figure 2 Results of titration of mouse anti-pn6B sera on immunotech plates coated with 2
 25 µg/ml of pn6B.

Figure 3 Results of titration of anti-pn14 sera on pn14 coated plates.

Description of the Invention

Soluble, intact recombinant CRM was first produced in protease-deficient *E.coli* (Bishai et.al 1987). However, the amount of protein production was very low. Subsequently, CRM₁₉₇
 30 was produced in *E.coli* cells as inclusion bodies (Stefan A, et al. J Biotechnol. Dec 20;156(4):245-52, 2010; International Application Publication No. WO 2011/126811, Chinese

Patent Application No. 200610042194) or as soluble protein directed to the periplasm by signal peptide (International Application Publication No. WO 2011/042516). The periplasm of *E.coli* is an oxidizing environment that allows the formation of disulfide bonds. CRM₁₉₇ has two disulfide bonds that are probably important for the correct folding and function, and for protein solubility.

It has been surprisingly discovered that a single, uncleaved chain of soluble recombinant CRM protein can be rapidly produced intracellularly and in commercial quantities from microorganisms and thereafter isolated and/or purified in large quantities and remain soluble. CRM is soluble in phosphate buffered saline (PBS, pH 7.5) and other similar buffers and can be concentrated to greater than 5 mg/ml in this and other buffers while remaining soluble. While CRM expressed in *Cornybacter* can be concentrated in these buffers, CRM made in *Pseudomonas* and expressed in the periplasm, cannot be easily concentrated in these same buffers (Pfenex Inc., San Diego, CA). A further advantage of intracellular expression compared with periplasmic expression is that greater expression levels are achieved because the periplasmic space is limiting compared to intracellular space.

Preferred CRM proteins produced are full length or partial regions such as, for example, peptides, single or multiple domains or epitopes, and any specific region expressed from native CRM coding sequences including CRM sequences that have been modified with one or more deletions, substitution and/or additions (e.g. conservative or non-conservative), and CRM sequences that have been modified with additional sequences (e.g., one or more promoters, start codons, and translation factor, ribosome or polymerase binding sites) that promote expression in a host organism. A preferred CRM protein is CRM₁₉₇. Preferred is expression of CRM protein that is soluble and not otherwise bound as insoluble inclusion bodies of the cell. Preferred expression systems for the expression and production of CRM proteins include microorganisms with an intracellular oxidative state. Preferred expression systems may be recombinant or native eukaryotic or prokaryotic cells wherein recombinant cells include cells that contain a non-native CRM coding sequence. Preferred prokaryotic cells are strains of *E.coli* or another bacterial strain that contains one or more genetic alterations (e.g., one or more deletions or mutations). Preferably the one or more genetic alterations shift the redox state of the cytoplasm of the cell to a more oxidative state, as compared to wild-type, for example as disclosed in U.S. Patent No. 7,410,788.

Alterations preferably reduce the activity of one

or more disulfide reductase genes and/or other genes that reduce the oxidative state of the cytoplasm. Preferably, reduced activity is due to non-expression or reduced expression of one, two or multiple disulfide reductase or other genes, or one or more mutations that reduce activity of one or more expressed disulfide reductase proteins or other proteins. Preferred strains of microbial cells (e.g., recombinant, engineered or native eukaryotic or prokaryotic cells) have increased abilities to produce natively folded proteins containing disulfide bonds yet remain as functional proteins. The method of the invention produces quantities of CRM proteins containing full, truncated or modified CRM amino acid sequences. Quantities of CRM protein produced according to the invention are surprising such as, for example, 600 mg or more of CRM protein per liter of bacterial cell culture.

One embodiment of the invention is directed to methods for the production of large quantities of CRM protein, and preferably CRM₁₉₇. Production quantities are typically quantified as mg/L of bacterial cell culture. CRM protein production, according to the methods of the invention, is 200mg/L or more, 300mg/L or more, 400mg/L or more, 500mg/L or more, 600 mg/L or more, 700 mg/L or more, 800 mg/L or more, 900 mg/L or more, 1,000 mg/L or more, 1,500mg/L or more, or 2,000mg/L or more. Preferred quantities CRM₁₉₇ of the invention includes related proteins containing full length and truncated CRM protein, as well as modified amino acid sequences of CRM protein. Modifications include one or more of conservative amino acid deletions, substitution and/or additions. A conservative modification is one that maintains the functional activity and/or immunogenicity of the molecule, although the activity and/or immunogenicity may be increased or decreased. Examples of conservative modifications of CRM include, but are not limited to amino acid modifications (e.g., single, double and otherwise short amino acid additions, deletions and/or substitutions), modifications outside of the 39 alpha-amino groups of lysine (primary amine groups of lysine) residues that are accessible for conjugation in forming a vaccine, modifications due to serotype variations of DT, modifications that increase immunogenicity or increase conjugation efficiency, modification that do not substantially alter binding to heparin, modifications that maintain proper folding or three dimensional structure, and/or modifications that do not significantly alter immunogenicity of the protein or the portions of the protein that provide protective immunity to DT.

Recombinant cells that are used in the method of the invention are preferably *E. coli* bacteria and, preferably, *E. coli* that are genetically engineered to shift the redox state of the

cytoplasm to a more oxidative state such as, for example, by mutation of one or more disulfide reductase genes such as, for example, an oxidoreductase, a dihydrofolate reductase, a thioredoxin reductase, a glutamate cysteine lyase, a disulfide reductase, a protein reductase, and/or a glutathione reductase. Preferably one or more disulfide reductase genes are mutated and rendered non-functional or marginally functional such that the redox state of the cytoplasm of the cell is shifted to a more oxidative state as compared to wild type. Oxidative protein folding involves the formation and isomerization of disulfide bridges and plays a key role in the stability and solubility of many proteins including CRM₁₉₇. Formation and the breakage of disulfide bridges is generally catalyzed by thiol-disulfide oxidoreductases. These enzymes are characterized by one or more Trx folds that consist of a four-stranded β -sheet surrounded by three α -helices, with a CXXC redox active-site motif. The assembly of various Trx modules has been used to build the different thiol oxidoreductases found in prokaryotic and in eukaryotic organisms. In the bacterial periplasm, the proteins are kept in the appropriate oxidation state by a combined action of the couples DsbB-DsbA and DsbD- DsbC/DsbE/DsbG (Inaba 2009, Gruber et al, 2006). Many protein expression systems are well known in the art and commercially available.

Especially preferred microbes include *E. coli* expression strains, for example, chemically competent *E. coli* K12 cells engineered to form disulfide bonded proteins in the cytoplasm (e.g., ORIGAMI™ (EMD Millipore) and SHUFFLE™ (New England Biolabs)). Other strains and types of cells and other *E. coli* strains with enhanced oxidative redox state also may be used. For example, ORIGAMI™ 2 host strains are K-12 derivatives that have mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which greatly enhance disulfide bond formation in the *E. coli* cytoplasm. These strains are kanamycin sensitive; like the original Origami strains, the *gor* mutation is still selected for by tetracycline. To reduce the possibility of disulfide bond formation between molecules, strains containing mutations in *trxB* and *gor* are recommended only for the expression of proteins that require disulfide bond formation for proper folding. SHUFFLE™ cells are chemically competent *E. coli* K12 cells engineered to form proteins containing disulfide bonds in the cytoplasm. Preferably these cells contain mutations in *trxB* and *gor* and cytoplasmic chaperon disulfide bond isomerase DsbC (*fhuA2 [lon] ompT ahpC gal* λ att::pNEB3-r1-*cDsbC* (SpecR, *lacI^q*) Δ *trxB* *sulA11* R(*mcr-73::miniTn10--Tet^S*)2 [*dcm*] R(*zgb-210::Tn10 --Tet^S*) *endA1* Δ (*mcrC-mrr*)114::IS10). Also

preferably, cells are suitable for T7 promoter driven protein expression and of the genotype F' *lac, pro, lacIQ / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt::pNEB3-r1-cDsbC (Spec^R, lacI^q) ΔtrxB rpsL150(Str^R) Δgor Δ(malF)3.*

SHUFFLE™ strains expresses constitutively a chromosomal copy of the disulfide bond

- 5 isomerase DsbC. DsbC promotes the correction of mis-oxidized proteins into their correct form. Cytoplasmic DsbC is also a chaperone that can assist in the folding of proteins that do not require disulfide bonds.

Bacterial cultures are preferably cultured at temperatures such that solubility of the expressed protein increases (e.g., CRM or CRM₁₉₇) as compared to solubility at higher
10 temperatures (e.g., 37°C). Preferred culture temperatures are 30°C or lower, preferably 25°C or lower, preferably 20°C or lower, preferably 18°C or lower, and preferably between 15°C and 32°C.

Another embodiment of the invention is directed to vectors for producing CRM and methods of producing all or a portion of a CRM protein, such as preferably CRM₁₉₇, soluble in
15 the cytoplasm of a cell and preferably a prokaryotic cell. Previous attempts to express CRM in E.coli intracellularly were based on monocistronic mRNA, encoding only the CRM sequence and resulted in inclusion body formation. Methods of producing soluble CRM in the cytoplasm of cells were developed using an expression vector that provides transcription of CRM in polycistronic mRNA. Polycistronic mRNA refers to messenger RNA that encodes two or more
20 polypeptides. In prokaryotic cell, genes that are involved in the same biochemical or physiological pathway are often grouped into an operon, controlling transcription of the genes into a single polycistronic mRNA. Genes (cistrons) in the operon are controlled by a ribosome binding site sequences and can be separated by a number of nucleotides or even overlapping sequences. For example, a stop codon of the first gene is downstream of the second gene start
25 codon, as in the galactose operon. Gene location in the operon has been shown to also strongly affect gene expression level via translational and mRNA stability effects (Smolke, C. D., and Keasling, J. D. (2002) Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. *Biotechnol. Bioeng.* 80, 762–76). The downstream gene expression level is found to be enhanced by the upstream gene expression via
30 translational coupling (Schümperli, D., McKenney, K., Sobieski, D. a, and Rosenberg, M. (1982) Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose

operon. Cell 30, 865–71). One preferred embodiment of the invention is a vector comprising a prokaryotic promoter and two cistrons encoding polypeptides, one of them being CRM. Each cistron comprises a ribosome binding site and an initiation codon such as, for example, ATG. The invention further includes inducing the expression vector to produce CRM protein and
5 isolating the CRM protein expressed. In one preferred embodiment, the first cistron preceding the CRM sequence contains the T7 tag sequence, overlapping with the CRM cistron, so that stop codon for the first cistron is downstream of the initiation codon of CRM (e.g. SEQ ID NO 15). The expression enhancer is further modified as SEQ ID NO 16 or SEQ ID NO 17. The first cistron preceding CRM coding sequence is termed an “expression enhancer sequence” (EES).
10 The expression vector contains (1) a promoter followed by a ribosome binding site and the expression enhancer sequence, and (2) a ribosome binding site and an ATG codon and the CRM coding sequence. The recombinant cell may be a prokaryotic or eukaryotic cell. Preferably the recombinant cell is a prokaryotic cell such as, for example, an *E. coli* cell or a derivative or strain of *E. coli*. Preferably, the recombinant cell modification comprises a reduced activity of one or
15 more disulfide reductase enzymes such as, for example, one or more of an oxidoreductase, a dihydrofolate reductase, a thioredoxin and a thioredoxin reductase, a protein reductase or a glutathione reductase. Preferably the reduced activity of the one or more disulfide reductase enzymes shifts the redox state of the cytoplasm of the recombinant cell to an oxidative state as compared with a non-recombinant cell. Preferably the CRM coding sequence encodes one or
20 more CRM epitopes, CRM peptide sequences, CRM domains, or combinations thereof. Preferably the CRM protein expressed by the cell is soluble and is intracellular, periplasmic or secreted. Preferably the recombinant cell is propagated at a temperature from about 15°C to about 32°C.

Another embodiment of the invention comprises recombinant cells such as, for example,
25 bacterial, mammalian or insect cells containing expressible CRM sequences and, preferably sequences of CRM₁₉₇. Preferred host cells include, but are not limited to, cells genetically engineered to shift the redox state of the cytoplasm to a more oxidative state. Preferred cells include prokaryotic or eukaryotic cells such as, for example, *E. coli* cell expression systems, *Baculovirus* Expression System and other bacterial and/or eukaryotic cellular expression
30 systems. Preferably the cells contain a protein expression system for expressing foreign or non-native sequences such as CRM peptides. Also preferable, the sequences to be expressed are

comprised of an expression vector which contains one or more of an inducible promoter (e.g., auto-inducible preferably with specific media), a start codon (e.g., ATG), a ribosome binding site, unspecified polypeptide sequence and CRM coding sequence transcribed into polycistronic mRNA. Also preferably, the expression vector contains a modified sequence between ribosome binding site and ATG starting codon, or between start codon and the sequence to be expressed. Preferred modified sequences or spacer sequences include, for example, a number of nucleotides more or less than 9 (e.g., between 7 and 12 nucleotides), and preferably not 9 nucleotides. Specific examples of spacer nucleotides that can be utilized in an expression system include but are not limited to GATATAC (SEQ ID NO 3), GATATACCA (SEQ ID NO 4), and GATATACCATAT (SEQ ID NO 5). Accordingly, another embodiment of the invention comprises an expression construction of CRM, nucleotide and amino acids sequences, with or without defined spacer sequences and with and without a host cell.

Another embodiment of the invention is directed to recombinant CRM₁₉₇ protein and the expression of recombinant CRM in *E. coli* or another host cell using an expression vector with an inducible promoter and/or a modified sequence between ribosome binding site and ATG starting codon. Preferably, the expression vector includes the lactose/IPTG inducible promoter, preferably a *tac* promoter, and the sequence between ribosome binding site and ATG starting codon. Preferably the expression system contains a spacer between the start codon and the expression sequence which is comprised of a number of nucleotides more or less than 9 (e.g., between 7 and 12 nucleotides), and preferably not 9 nucleotides. Specific examples of spacer nucleotides that can be utilized in an expression system include but is not limited to those identified herein. It was surprisingly discovered that the use of spacers of length seven or twelve resulted in dramatically increased levels of CRM₁₉₇ expression when compared to spacers of nine nucleotides.

Another embodiment of the invention comprises an expression construction of CRM, nucleotide and amino acids sequences, with or without defined spacer sequences, as disclosed herein, and with or without an enhancer region. Enhancers regions promote expression of the downstream CRM sequence by translational coupling enhancing correct folding of CRM resulting in protein solubility. Enhancers regions also promote protein expression by adding one or more sequences that promote nucleic acid recognition for increased expression (e.g., start codon, enzyme binding site, translation or transcription factor binding site). Preferably, an

enhancer of the invention contain a ribosome binding site with a start codon upstream of and with a coding sequence that differs from the coding sequence of the CRM protein.

Another embodiment of the invention is directed to recombinant CRM, and in particular CRM₁₉₇, purified according to the methods of invention. Purification preferably comprises heparin or heparin-like affinity chromatography. It was surprisingly discovered that CRM₁₉₇ contains the sequence-based motif of typical heparin binding sites XBBXBX (SEQ ID NO 6) where B is a lysine or arginine and X a hydrophobic residue (Cardin AD, Weintraub HJ., 1989: Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* 9: 21–32). This motif is located in the CRM₁₉₇ receptor-binding domain and comprises of the following amino acids: GRKIRMRCR (SEQ ID NO 7), where G (Glycine), I (Isoleucine), M (Methionine) and C (Cysteine) are hydrophobic residues. Presence of heparin binding site allows the use of heparin or heparin-like resins in the purification. Heparin-like resins include resins containing functional sulfate groups, such as dextran sulfate, e.g. Dextran sulfate (Sterogene), Capto Devirs (GE) or sulfate esters, e.g. Cellufine Sulfate (Asahi Kasei Bioprocess).

In a first step, crude *E. coli* extract may be clarified, for example, preferably by centrifugation or depth filtration. Optionally cleared lysate may be fractionated further, preferably by adding salts that have effect on protein solubility and salting out CRM₁₉₇. In the second step, clarified lysate or re-solubilized salted out fraction containing CRM₁₉₇ may be applied, for example, to anion exchange resin under conditions when CRM₁₉₇ is in flow through.

In the third step, the flow through fraction containing CRM₁₉₇ may be applied to a column. Preferred column resins include, but are not limited to dextran sulfate resins, CELLUFINE™ resins (Chisso Corporation; chromatography gel), active sulfated resins, phosphate resins, or heparin or heparin-like resins. Preferably binding of CRM to resin is performed in a low salt buffer and eluted in higher salt buffer, yielding highly purified CRM₁₉₇. Preferred binding buffers contain, for example, one or more chaotropic agents, NaCl, KCl, glycerol, isopropyl alcohol, ethanol, arginine, acetate, guanidine, urea, ATP, one or more mono-, di-, tri-, and/or poly-phosphates, sulfates or pyrophosphates, and combinations thereof. Preferred elution buffers contain, for example, higher concentration of one or more components of the binding buffer.

Other preferred purification methods include any one or combination of an anion exchange chromatography, hydrophobic interaction chromatography and/or Cibacron-Blue resin (CN 101265288A, U.S. Patent No. 8,383,783). Purification method of the invention produce

recombinant CRM protein (e.g., CRM₁₉₇) at high yields, as discussed herein, and with a purity level of greater than 80%, preferably greater than 85%, preferably greater than 90%, preferably greater than 95%, preferably greater than 99%, and preferably with an even greater purity.

Another embodiment of the invention is directed to methods to characterize recombinant DT and CRM proteins (e.g., binding activity) and, in particular CRM₁₉₇, which contain a receptor binding domain (see SEQ ID NO 2). These methods comprise determination of the binding activity of proteins containing native or modified sequence of receptor binding domain of DT. Such modifications preferably preserve the ability of CRM to bind to HB-EGF (heparin binding epidermal growth factor). The method is applicable to both crude and purified CRM₁₉₇.

Binding activity represents binding to the soluble form of diphtheria toxin receptor HB-EGF (DTR). These methods comprise, preferably, determining the binding CRM₁₉₇ to DTR and detection of with molecules (e.g., antibodies, antibody fragments, antigens) specific to the properly folded structure, the complex, binding, and/or the binding sites, and preferably in an ELISA format. Assays to determine and quantitate binding allow for the rapid determination that CRM₁₉₇ is correctly folded, as only properly folded CRM₁₉₇ binds to the receptor. Thus, the method monitors correct folding of manufactured CRM₁₉₇ and related proteins during the development, production and purification process. In addition, this characterization method can be used to identify and track CRM protein after conjugation with another molecule such as in vaccine production. Using the detection method of the invention, properly folded and configured conjugated CRM protein can be monitored during the development of a vaccine for the treatment and/or prevention of diseases and disorders in patients.

Another embodiment of the invention comprises methods for conjugating CRM protein for vaccine production, such as, for example, by conjugation with a polysaccharide. Also included are the conjugation of proteins, peptides oligosaccharides and haptens. Another embodiment of the invention comprises a vaccine containing CRM protein of the invention. Another embodiment of the invention is directed to CRM protein of the invention fused genetically or chemically with another molecule, such as another protein or polysaccharide. Fusion is preferably by one or more covalent bonds between the molecules.

The following examples illustrate embodiments of the invention, but should not be viewed as limiting the scope of the invention.

Examples

Example 1. CRM₁₉₇ expression detected from expression vectors containing 7, 8, 9 or 12 nucleotides sequence between RBS and initial ATG codon.

DNA encoded mature CRM₁₉₇ was cloned into expression vector in polycistronic format resulting in the following sequence of DNA regulatory and coding fragments: tac promoter-
 5 ribosome binding site-ATG codon-T7 tag-ribosome binding site-ATG codon-CRM coding sequence-stop codon. Different E coli strains, including ORIGAMI 2, C41, were tested as expression strains (Figure 1).

Example 2. CRM₁₉₇ expresses soluble in Origami 2, *E. coli* expression strain that allows formation of disulfide bonds in the cytoplasm.

10 CRM₁₉₇ was expressed insoluble at 37°C. When expression temperature was dropped below 37°, solubility of the protein expressed in ORIGAMI™ 2 cells and SHUFFLE™ cells, but not in the other tested E. coli strains, increases. CRM₁₉₇ is mostly soluble when expressed in ORIGAMI™ 2 cells at 18°C.

Example 3. Expression enhancer sequence (EES) in CRM₁₉₇ expression.

15 The EES promotes transcription of CRM sequence in a CRM-containing vector and results in polycistronic mRNA that translates into two proteins; a short EES peptide and a CRM peptide. The coding sequence of the native CRM gene was analyzed for potential 3D structure formation and found to contain a number of potential hairpins, which could inhibit translation. A CRM sequence was created that would potentially result in an mRNA with no hairpins structures
 20 yet translate the same CRM amino acid sequence. This gene sequence is referred to an optimized CRM sequence and comprises SEQ ID NO 8.

The optimized CRM sequence expresses well in both *E. coli* (e.g., BL21) and in *E. coli* engineered to contain an oxidized cytoplasm (e.g., Shuffle). CRM peptide translated from polycistronic mRNA produces a full length protein and is believed to be more stable than the
 25 native CRM coding sequence. Unlike the native CRM sequence, the optimized CRM sequence expressed as full-length and as a soluble protein in Shuffle cells. In addition, compared to native CRM, higher expression of the optimized CRM sequence is observed with a lower cell density and with increased binding to chromatography resin resulting in greater production levels of CRM protein.

30 **Example 4. Ammonium sulfate precipitation of CRM₁₉₇ from cell lysate.**

SHUFFLE™ cells expressing CRM₁₉₇ were open using microfluidizer and 1M of sodium chloride was added to the cell lysate. To this was added enough ammonium sulfate to equal 1M followed by centrifugation for 30 minutes at 20,000x g, which removed mis-folded CRM₁₉₇ and most of the bacterial proteins. Following clarification the ammonium sulfate concentration was further increased to 2.2M. The precipitate, which is mainly CRM₁₉₇ was collected and re-solubilized in a low conductivity buffer.

Example 5. Purification of CRM₁₉₇ on a Heparin column.

Ammonium sulfate precipitated CRM₁₉₇ was resolubilized in 20mM Tris-HCl pH 8.0 to achieve conductivity 5 mS/cm and loaded on a column containing Heparin Sepharose™ CL-6B resin (GE). The purification was performed under the following conditions: flow rate was 5ml/min, wash buffer A: 20mM Tris-HCl pH8. Elution was done with a buffer B 0-100% gradient, buffer B: buffer A + 1M NaCl in 20 CV. Eluted CRM₁₉₇ was analyzed by SDS-PAGE in reduced and non-reduced conditions. The purity of eluted CRM₁₉₇ was greater than 95%. The protein reduced with DTT appears as a single polypeptide confirming that the intact form of CRM₁₉₇ is expressed in *E. coli*.

Example 6. Purification of CRM₁₉₇ on Capto Devirs column.

SHUFFLE™ cells expressing CRM₁₉₇ were opened using a microfluidizer in 1xPBS, pH 7.4, 1% sodium pyrophosphate. The lysate was clarified using depth filtration. Clarified lysate was loaded on a column containing Q Sepharose™ XL (GE) and flow through fraction was collected. To reduce volume and conductivity flow through fraction was subjected to tangential flow filtration using 10K cassette (Sartorius). Capto Devirs resin was equilibrated with 25mM sodium phosphate buffer, pH8.0. CRM₁₉₇ was bound to the column under the following conditions: conductivity was less than 10mS/cm, in a binding buffer containing a chaotropic agent (e.g., in this case urea), wash buffer was 25mM sodium phosphate, pH8.0. Elution was done with NaCl. Eluted CRM₁₉₇ was analyzed by SDS-PAGE under reduced and non-reduced conditions. The purity of eluted CRM₁₉₇ was greater than 95%. The protein, reduced with DTT, appears as a single polypeptide confirming that CRM₁₉₇ remains intact during purification process.

Example 7. Binding assay for the CRM₁₉₇ characterization.

The recombinant soluble diphtheria toxin receptor HB-EGF (DTR) (Sigma) was bound to the ELISA plate. Blocking solution of 5% dry non-fat milk was used to prevent high

background. Recombinant CRM₁₉₇ diluted in 1xPBS, pH7.4, 0.1% Twin 20 was incubated on the plate for 1 hour at 37°C. CRM₁₉₇ bound to HB-EGF was detected by rabbit polyclonal anti-CRM₁₉₇ antibody and goat anti-rabbit antibody conjugated to soybean peroxidase (Fina BioSolutions; Rockville, MD). Denatured recombinant CRM₁₉₇ did not bind to the receptor.

5 **Example 8. CRM₁₉₇ produced in *E. coli* binds to DTR similarly to CRM from *Corynebacterium* and *Pseudomonas*.**

ELISA plates were coated with soluble HB-EGF (heparin-binding EGF-like growth factor) and blocked with 5% dry non-fat milk. CRM₁₉₇ was bound to the receptor and detected with rabbit anti-CRM₁₉₇ polyclonal antibody and goat anti-rabbit polyclonal conjugated with
10 SBP. CRM₁₉₇ expressed in *E. coli* showed the same affinity to HB-EGF as CRM produced in *Corynebacterium* and *Pseudomonas*.

Example 9 CRM₁₉₇ is a carrier protein

CRM₁₉₇ was expressed and purified according to the method of this invention (Example 1) and chemically linked (conjugated) to pneumococcal capsule polysaccharides serotypes 14
15 and 6B using CDAP chemistry (Lees, A., Producing immunogenic constructs using soluble carbohydrates activated via organic cyanating reagents. See U.S. Patent Nos. 5,651,971; 5,693,326 and 5,849,301). The conjugates were purified from unconjugated protein and polysaccharide. BALB/c female mice were immunized subcutaneously with the conjugate according to the schedule in Table 1. Mice were immunized in complete Freund's adjuvant and
20 boosted twice in incomplete Freund's adjuvant and day 57 bleeds were taken.

Table 1

Serotype	Primary in CFA*	Boost IFA**	Boost IFA**	D57
6B	20 ug	10 ug day 28	10 ug day 48	bleed
14	20 ug	5 ug day 21	5 ug day 48	bleed

*60% Complete Freund's Adjuvant; **60% Incomplete Freund's Adjuvant

Sera was tested for reactivity by ELISA on a Brandtech Immunograde plate coated with 2 µg/ml of Pn6B or Pn14 (from ATCC) using gamma-specific detection. Results in Figure 2
25 show a strong reactivity with Pn6B. Mouse 5086 was used for hybridoma production and three of the resulting hybridomas were used to prepare highly specific mouse anti-6B monoclonal antibodies. The results of the serum titration against Pn14 coated plates are shown in Figure 3. Mouse 1397 was subsequently used for the production of four highly specific mouse monoclonal

antibodies reactive with P14 polysaccharide. Unconjugated polysaccharide does not give a significant ELISA absorbance.

Other embodiments and uses of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein.

5

The term comprising, where ever used, is intended to include the terms consisting and consisting essentially of. Furthermore, the terms comprising, including, containing and the like are not intended to be limiting. It is intended that the specification and examples be considered

10 exemplary only with the true scope and spirit of the invention indicated by the following claims.

Sequences

SEQ ID NO 1 CRM197

15 GADDVVDSSK SFVMENFSSY HGTPKGYVDS IQKGIQPKS GTQGNYYYYDW
KEYFSTDNKY DAAGYSVDNE NPLSGKAGGV VKVTYPGLTK VLALKVDNAE
TIKKELGLSL TEPLMEQVGT EEFIKRFG DG ASRVVLSLPF AEGSSSVEYI
NNWEQAKALS VELEINFE TR GKRQGDAMYE YMAQACAGNR VRRSVGSSLS
CINLDWDVIR DTKTKIESL KEHGPIKNKM SESPNTVSE EKAKQYLE EF
20 HQTALHEPEL SELKTVTGTN PVFAGANYAA WAVNVAQVID SETADNLEKT
TAALSILPGI GSVMGADGA VHHNTEEIVA QSIALSSLMV AQAIPLVGEL
VDIGFAAYNF VESIINLF QV VHNSYNRPAY SPGHKTQPFL HDGYAVSWNT VEDSIIRT
GF QGESGHDIKI TAENTPLPIA GVLLPTIPGK LDVNKSKT HI SVNGRKIRM
CRAIDGDVTF CRPKSPVYVG NGVHANLH VA FHRSSSEKIH SNEISSDSIG
VLGYQKTVDH TKVNSKLS LF FEIKS

25

SEQ ID NO 2 Domain of CRM197

30 SPGHKTQPFL HDGYAVSWNT VEDSIIRT GF QGESGHDIKI TAENTPLPIA
GVLLPTIPGK LDVNKSKT HI SVNGRKIRM CRAIDGDVTF CRPKSPVYVG
NGVHANLH VA FHRSSSEKIH SNEISSDSIG VLGYQKTVDH TKVNSKLS LF FEIKS

SEQ ID NO 3 GATATAC spacer

SEQ ID NO 4 GATATACCA spacer

SEQ ID NO 5 GATATACCATAT spacer

35 **SEQ ID NO 6** XBBXB putative heparin binding site

SEQ ID NO 7 GRKIRMRCR heparin binding site

SEQ ID NO 8 Optimized CRM sequence

40 ATGGGTGCTGATGATGTTGTTGATTCTCTAAGTCTTTCGTGATGGAAAATTTCTCGT
CCTATCACGGTACCAAGCCTGGCTATGTGGATAGCATTCAAAAGGGTATTCAAAAAC

CGAAGTCTGGTACCCAGGGCAACTACGATGACGATTGGAAAGAGTTTTACAGCACC
 GACAACAAATATGACGCGGCAGGCTACAGCGTTGATAATGAAAATCCGCTGAGCGG
 TAAGGCTGGCGGCGTCTGTTAAGGTTACCTATCCGGGTCTGACGAAAGTGCTGGCCCT
 GAAAGTTGACAATGCTGAAACCATCAAAAAAGAACTGGGTCTGAGCTTGACCGAGC
 5 CGCTGATGGAACAGGTTGGTACTGAAGAATTCATTAAACGTTTTGGTGACGGCGCGA
 GCCGTGTTGTGCTGTCCCTGCCGTTTGCCGAGGGTTCTAGCTCCGTGGAGTATATCA
 ACAATTGGGAACAGGCGAAAGCGTTGAGCGTCGAGCTGGAAATCAATTTGAGACT
 CGTGGTAAGCGTGCCCAAGATGCGATGTACGAGTACATGGCCCAGGCATGTGCGGG
 TAACCGCGTCCGTCCGACGCTCGGCAGCTCCCTGAGCTGCATTAACCTGGACTGGGA
 10 CGTGATCCGCGACAAGACTAAGACCAAGATTGAGAGCCTGAAAGAGCACGGTCCGA
 TTAAGAACAAAATGTCCGAGTCTCCGAACAAAACGGTGAGCGAAGAAAAAGCCAA
 ACAGTATCTGGAAGAATTCCATCAGACCGCCCTGGAGCACCCAGAGCTGAGCGAGC
 TGAAAACCGTCACCGGCACGAATCCGGTTTTTGGCGGTGCGAACTACGCGGCATGG
 GCAGTCAATGTTGCGCAAGTCATCGACAGCGAAACGGCTGATAACTTGGAGAAAAC
 15 CACCGCGGCACTGAGCATTCTGCCGGGCATCGGTAGCGTTATGGGCATTGCGGACG
 GTGCCGTGCATCACAATACCGAAGAAATTGTGCGCGAGAGCATCGCATTGTCTAGCC
 TGATGGTTGCACAGGCCATTCCGCTGGTAGGCGAATTGGTGGATATCGGTTTCGCGG
 CTTACAATTTCTGTTGAGTCGATCATTAACCTGTTTCAAGTCGTTTACAATAGCTATAA
 CCGTCCGGCATAACAGCCCGGGTCATAAGACGCAACCGTTTCTGCATGATGGCTATGC
 20 CGTGAGCTGGAACACGGTCGAGGATTGATTATCCGTACCGGTTTTTCAGGGTGAGAG
 CGGTCACGACATCAAAATCACCGCGGAGAACACGCCGCTGCCTATTGCGGGCGTCC
 TGCTGCCGACGATCCCGGGCAAACCTGGACGTTAACAAGAGCAAGACCCATATCAGC
 GTCAACGGTCGTAAGATTTCGCATGCGTTGTCTGCAATCGACGGTGACGTGACGTTT
 TGCCGCCCAAAAAGCCCGGTGTACGTGGGTAAACGGCGTGACGCGAATCTGCATGT
 25 CGCGTTCCACCGCTCCTCAAGCGAGAAAATCCACAGCAATGAAATTAGCAGCGACA
 GCATTGGTGTGTTGGGCTACCAAAAGACCGTGGATCACACCAAGGTTAATAGCAAG
 CTGAGCCTGTTCTTTGAGATCAAAAGC

30 **SEQ ID NO 9 Not optimized CRM sequence**

ATGGGTGCCGATGACGTGGTTGACTCTTCCAAAAGCTTCGTCATGGAAAACCTTCAGC
 TCCTATCACGGCACTAAACCGGGTTATGTGACAGCATCCAGAAAGGCATCCAGAA
 ACCGAAATCTGGCACTCAGGGTAACCTATGACGACGACTGGAAAGAGTTCTACTCTA
 35 CCGACAACAAATACGACGCGGCTGGTTATTCTGTGGACAACGAAAACCCGCTGTCT
 GGTAAGCTGGTGGTGTGTTAAAGTGACCTACCCGGGTCTGACCAAAGTTCTGGCT
 CTGAAAGTGGAACGCCGAAACCATCAAAAAAGAACTGGGTCTGTCTCTGACCGA
 ACCGCTGATGGAACAGGTAGGTACCGAGGAATTCATCAAACGTTTTGGTGATGGTG
 CGTCCCGTGTGTTGTTACTGTCTCTGCCATTTGCCGAAGGTTCTAGCTCTGTGAGTACAT
 40 CAACAACCTGGGAGCAGGCCAAAGCTCTGTCTGTGGAACCTGGAAATCAACTTCGAGA
 CCCGTGGTAAACGTGGTCAGGACGCAATGTATGAATACATGGCACAGGCTTGCGCG
 GGTAACCGTGTACGTCGTTCTGTAGGTTCTTCCCTGTCTTGTCATCAACCTGGACTGGG
 ATGTCATCCGTGACAAAACCAAAACCAAAATCGAGTCCCTGAAAGAGCACGGTCCG
 ATCAAAAACAAAATGAGCGAATCTCCGAACAAAACGGTCTCTGAGGAAAAAGCGA
 45 AACAGTACCTGGAAGAATTCCATCAGACCGCCCTGGAACACCCGGAACGTGTCTGAA
 CTGAAAACCGTTACCGGTACTAACCCGGTTTTTCGCAGGTGCTAACTACGCAGCGTGG

GCGGTTAACGTAGCCCAGGTAATCGATTCCGAAACCGCAGACAACCTGGAAAAAAC
 GACTGCGGCTCTGTCTATTCTGCCGGGTATTGGTAGCGTGATGGGTATTGCAGATGG
 TGCAGTTCACCACAACACGGAAGAAATCGTTGCGCAGTCTATCGCTCTGTCTTCTCT
 GATGGTAGCACAGGCGATCCCGCTGGTTGGTGAACCTGGTTGACATTGGCTTCGCGGC
 5 CTACAACTTCGTTGAATCCATCATCAACCTGTTCCAGGTTGTGCACAACTCTTACAAC
 CGTCCAGCTTACTCTCCGGGTCACAAAACCCAGCCGTTCTTGCACGACGGTTATGCG
 GTTTCTTGGAACACCGTTGAAGACAGCATCATCCGTAAGTTTCCAGGGTGAATCT
 GGCCACGACATCAAAATCACTGCTGAAAACACCCCGCTGCCGATCGCAGGTGTTCTC
 CTGCCAACTATTCCGGGTAAACTGGACGTGAACAAATCCAAAACGCACATCTCCGT
 10 GAACGGTCGTAAAATCCGCATGCGTTGTCGTGCGATTGATGGTGACGTTACTTTCTG
 TCGTCCGAAATCTCCGGTCTACGTAGGTAACGGTGTACATGCTAACCTCCATGTAGC
 GTTCCACCGTTCTTCTTCCGAGAAAATCCACTCCAACGAGATCTCTAGCGACTCTAT
 CCGTGTTCTGGGTTACCAGAAAACCGTTGACCACACCAAAGTGAACCTCCAACTCA
 GCCTGTTCTTCGAAATCAAATCT

15

SEQ ID NO 10 crm 7
 GAGCTCTAAGAAGGAGATATACATGGGTGCCGATGACGTGGTTGACTCT

SEQ ID NO 11 crm 7_2
 20 GAGCTCTTAAGAAGGAGATATACATGGGTGCCGATGACGTGGTTGACTCT

SEQ ID NO 12 crm 8
 GAGCTCTAAGAAGGAGATATACAATGGGTGCCGATGACGTGGTTGACTCT

25 SEQ ID NO 13 crm 9
 GAGCTCTAAGAAGGAGATATACACATGGGTGCCGATGACGTGGTTGACTCT

SEQ ID NO 14 crm 12
 GAGCTCTAAGAAGGAGATATACCATATATGGGTGCCGATGACGTGGTTGACTCT

30 SEQ ID NO 15

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTAGCATGACT
 GGTGGACAGCAAATGGGTCTGGGATCCGAATTCGAGCTCTAAGAAGGAGATATACC

SEQ ID NO 16

35 TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTAGCATGACT
 GGTAAGGAGATATACC

SEQ ID NO 17

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTAGCATGACT
 GGTGCGMAYCCATTCAGTGAAGAAGRAGSTTYATT

THE EMBODIMENTS OF THE INVENTION FOR WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE ARE CLAIMED ARE DEFINED AS FOLLOWS:

1. A method of producing all or a portion of a diphtheria cross reacting material (CRM) protein comprising:

providing a recombinant cell that contains an expression vector, wherein the recombinant cell has a reduced activity of one or more disulfide reductase enzymes and contains an inducible promoter functionally linked to a polycistronic genetic sequence wherein at least one cistron encodes the CRM protein;

inducing the expression vector to produce CRM protein cytoplasmically; and

isolating the CRM protein expressed which is mostly soluble CRM protein.

2. The method of claim 1, wherein each cistron contains a ribosome binding site and an initiation codon.

3. The method of claim 1, wherein the polycistronic genetic sequence contains at least one spacer between one or more ribosome binding sites and one or more initiation codons.

4. The method of claim 1, wherein the recombinant cell is propagated at a temperature from about 15°C to about 32°C.

5. The method of claim 1, wherein the CRM protein is isolated from the cell by chromatography.

6. The method of claim 5, wherein the chromatography comprises a dextran sulfate resin, a gel resin, an active sulfated resin, a phosphate resin, a heparin resin or a heparin-like resin.

7. Diphtheria cross reacting material (CRM) protein produced by the method of claim 5, wherein the recombinant cell is *E. coli* and the CRM protein is soluble, has no leader sequence and is tagless.

8. A method of producing all or a portion of a diphtheria cross reacting material (CRM) protein comprising:

providing a recombinant cell that contains an expression vector, wherein the recombinant cell has been modified to shift the redox status of the cytoplasm to a more oxidative state as compared to an unmodified recombinant cell and the expression vector contains an inducible promoter functionally linked to a CRM coding sequence, a spacer sequence between a

ribosome binding site and a start codon, and an expression enhancer region upstream of the CRM coding sequence;

inducing the expression vector to produce CRM protein cytoplasmically; and

isolating the CRM protein expressed, wherein the CRM protein is mostly soluble.

9. The method of claim 8, wherein the recombinant cell is an *E. coli* cell or a derivative or strain of *E. coli*.

10. The method of claim 8, wherein the recombinant cell modification comprises a reduced activity of one or more disulfide reductase enzymes.

11. The method of claim 10, wherein the one or more disulfide reductase enzymes comprises one or more of an oxidoreductase, a dihydrofolate reductase, a thioredoxin reductase, a protein reductase or a glutathione reductase.

12. The method of claim 8, wherein reduced activity of the one or more disulfide reductase enzymes shifts the redox state of the cytoplasm of the recombinant cell to an oxidative state as compared with a non-recombinant cell.

13. The method of claim 8, wherein the CRM coding sequence encodes one or more CRM epitopes, CRM peptide sequences, CRM domains, or combinations thereof.

14. The method of claim 8, wherein the CRM coding sequence encodes CRM₁₉₇.

15. The method of claim 8, wherein the spacer comprises between 5 and 20 nucleotides.

16. The method of claim 15, wherein the spacer comprises 9 nucleotides.

17. The method of claim 8, wherein the expression enhancer comprises a ribosome binding site upstream of the CRM coding sequence and an ATG codon.

18. The method of claim 8, wherein the recombinant cell is propagated at a temperature from about 15°C to about 32°C.

19. The method of claim 8, wherein the CRM protein is isolated from the cell by chromatography.

20. The method of claim 19, wherein the chromatography comprises a dextran sulfate resin, a gel resin, an active sulfated resin, a phosphate resin, a heparin resin or a heparin-like resin.

21. Diphtheria cross reacting materials (CRM) protein produced by the method of claim 19, wherein the recombinant cell is *E. coli* and the CRM protein is soluble, has no leader sequence and is tagless.

22. The method of claim 8, further comprising conjugating the isolated CRM protein to an antigen.

23. The method of claim 22, wherein the conjugated CRM protein is a vaccine.

24. A diphtheria cross reacting material (CRM) protein vaccine produced by the method of claim 23, wherein the recombinant cell is *E. coli* and the CRM protein is soluble, has no leader sequence and is tagless.

25. A method of producing all or a portion of a diphtheria cross reacting material (CRM) protein comprising;

providing a recombinant cell that contains an expression vector, wherein the recombinant cell has been modified to shift the redox status of the cytoplasm to a more oxidative state as compared to an unmodified recombinant cell and the expression vector contains a promoter functionally linked to a CRM coding sequence;

expressing CRM protein from the CRM coding sequence cytoplasmically; and

isolating the CRM protein expressed which is mostly soluble CRM protein.

26. The method of claim 25, wherein the recombinant cell is a prokaryotic or a eukaryotic cell.

27. The method of claim 26, wherein the prokaryotic cell is an *E. coli* cell or a derivative or strain of *E. coli*.

28. The method of claim 25, wherein the promoter is constitutive or inducible.

29. The method of claim 25, wherein the CRM coding sequence encodes one or more CRM epitopes, CRM peptide sequences, CRM domains, or combinations thereof.

30. The method of claim 25, wherein the CRM coding sequence encodes CRM₁₉₇.

31. The method of claim 25, wherein the modified recombinant cell has reduced activity of one or more disulfide reductase enzymes.

32. The method of claim 31, wherein the one or more disulfide reductase enzymes comprises one or more of an oxidoreductase, a dihydrofolate reductase, a thioredoxin, a thioredoxin reductase, a protein reductase or a glutathione reductase.

33. The method of claim 25, wherein the expression vector contains a spacer sequence between a ribosome binding site and an initiation codon.

34. The method of claim 33, wherein the spacer comprises between 5 and 20 nucleotides.

35. The method of claim 34, wherein the spacer comprises 9 nucleotides.

36. The method of claim 25, wherein the expression vector contains an expression enhancer.
37. The method of claim 36, wherein the expression enhancer comprises a ribosome binding site upstream of the CRM coding sequence and an initiation codon.
38. The method of claim 25, wherein isolating comprises:
- loading the CRM protein, which comprises mostly soluble CRM protein, onto a chromatography column containing a resin with a loading buffer;
 - washing the resin with one or more washing buffers;
 - eluting CRM protein from the resin with an elution buffer.
39. The method of claim 38, wherein the loading buffer and the washing buffer may be the same.
40. The method of claim 38, wherein the loading buffer and the one or more washing buffers are low conductivity buffers that have a conductivity of less than 10 mS/cm.
41. The method of claim 38, wherein the elution buffer is a high conductivity buffer with a conductivity of more than 10 mS/cm.
42. The method of claim 38, wherein the resin is selected from the group consisting of a dextran sulfate resin, a gel resin, an active sulfated resin, a phosphate resin, a heparin resin and a heparin-like resin.
43. A method of characterizing folding of diphtheria toxin or diphtheria cross reacting materials (CRM) protein comprising:
- contacting soluble diphtheria toxin or soluble CRM protein to Heparin-binding Epidermal Growth factor-like growth factor (HB-EGF);
 - determining the amount of binding of diphtheria toxin or CRM protein to HB-EGF;
- and
- determining the folding of diphtheria toxin or CRM protein by the amount of binding determined, wherein binding indicates correct folding.
44. The method of claim 43, wherein the diphtheria toxin or CRM contains a receptor binding domain.
45. The method of characterizing folding of CRM protein according to claim 43, wherein the CRM protein comprises CRM₁₉₇.
46. The method of claim 43, wherein at least one of the diphtheria toxin or CRM protein and the HB-EGF is bound to a solid support.

47. The method of claim 43, wherein the amount of binding of diphtheria toxin or CRM protein to HB-EGF is determined by an ELISA.

48. The method of claim 43, wherein CRM protein that binds to HB-EGF is soluble in PBS and concentrated to 5 mg/ml or greater in a buffer of about pH 7.5.

49. An expression vector comprising a promoter and two or more cistrons each encoding a protein, wherein at least one cistron encodes diphtheria cross reacting material (CRM) protein for cytoplasmic expression with no leader sequence, and each cistron has a ribosome binding site and an initiation codon.

50. The expression vector of claim 49, further comprising a spacer between the ribosome binding site and the initiation codon.

51. The expression vector of claim 50, wherein the spacer comprises from 5 to 20 nucleotides.

52. The expression vector of claim 51, wherein the spacer comprises 9 nucleotides.

53. A method of producing all or a portion of a diphtheria cross reacting material (CRM) protein comprising:

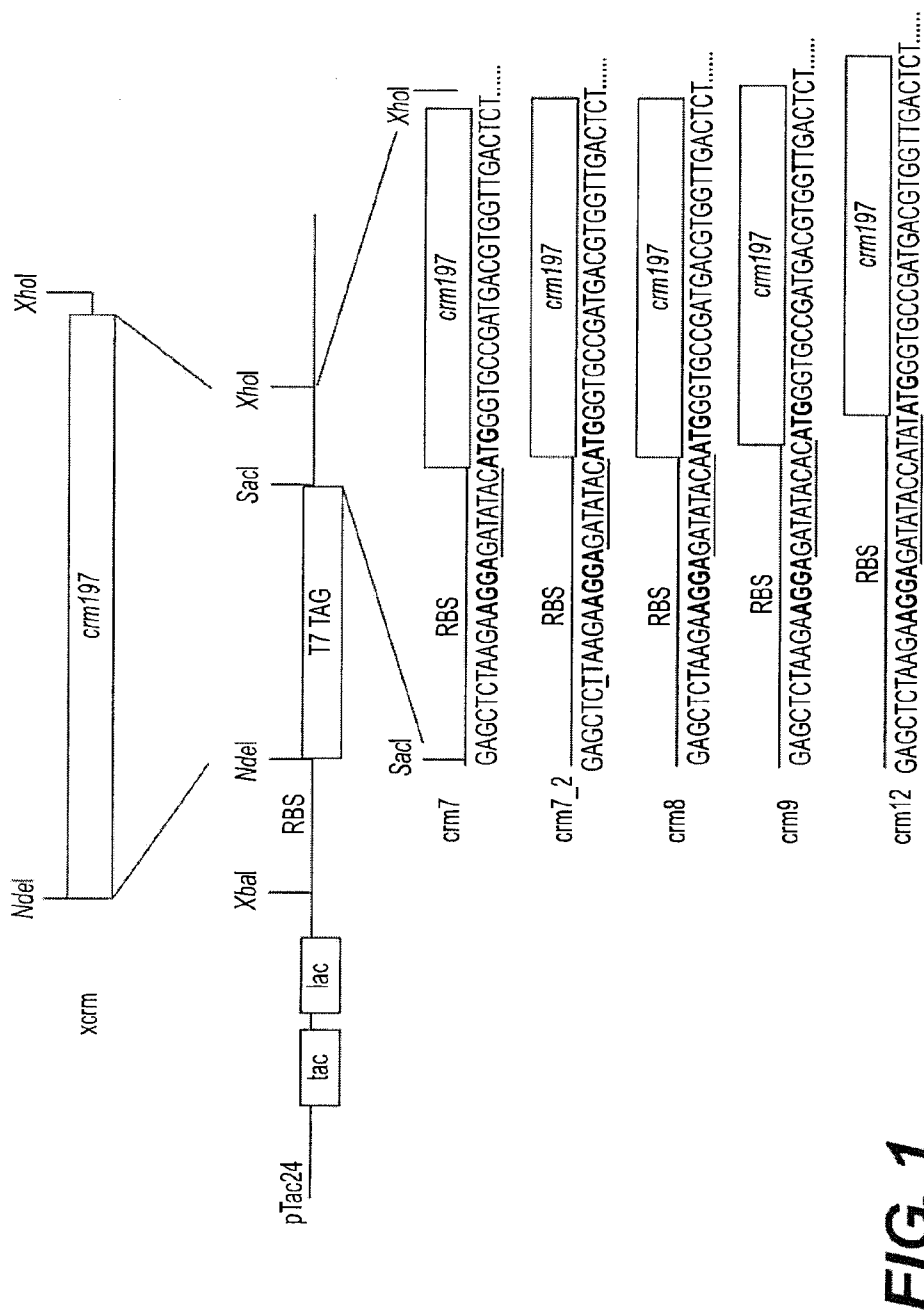
providing a recombinant cell that contains an expression vector, wherein the recombinant cell has a reduced activity of one or more disulfide reductase enzymes and contains a promoter functionally linked to a polycistronic genetic sequence wherein at least one cistron encodes the CRM protein; and

isolating the CRM protein expressed which is mostly soluble CRM protein.

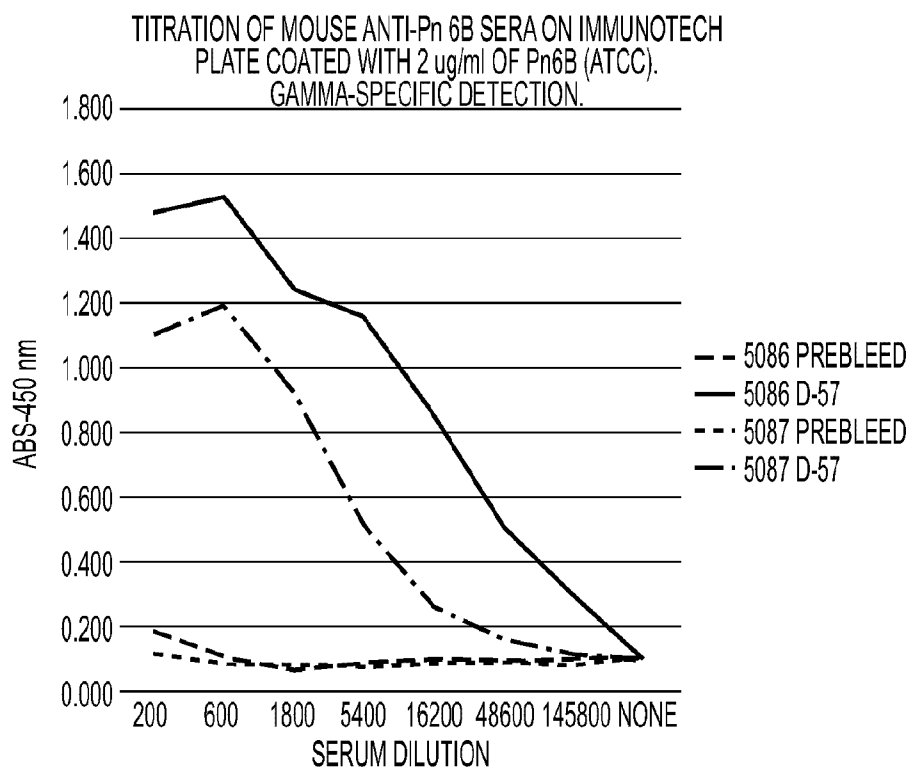
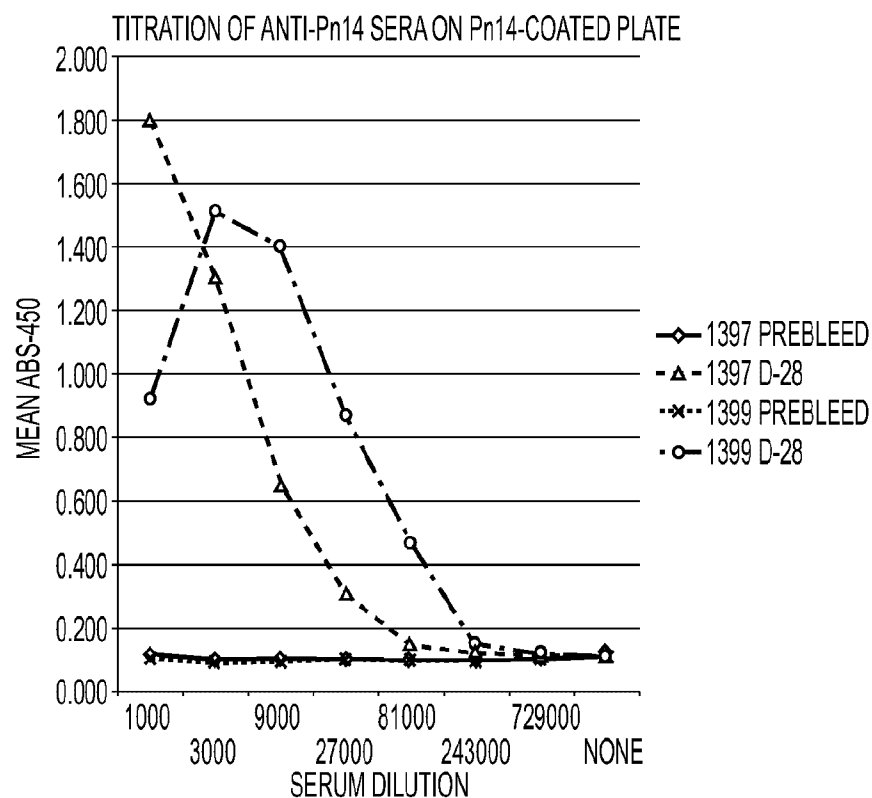
54. A method of producing all or a portion of a diphtheria cross reacting material (CRM) protein comprising:

providing a recombinant cell that contains an expression vector, wherein the recombinant cell has been modified to shift the redox status of the cytoplasm to a more oxidative state as compared to an unmodified recombinant cell and the expression vector contains a promoter functionally linked to a CRM coding sequence, a spacer sequence between a ribosome binding site and a start codon, and an expression enhancer region upstream of the CRM coding sequence; and

isolating the CRM protein expressed, wherein the CRM protein is mostly soluble.

**FIG. 1**

2/2

**FIG. 2****FIG. 3**

