



US 20230183307A9

(19) **United States**
(12) **Patent Application Publication**
Bishai et al.

(10) **Pub. No.: US 2023/0183307 A9**
(48) **Pub. Date: Jun. 15, 2023**
CORRECTED PUBLICATION

(54) **METHODS OF PRODUCING
AGGREGATE-FREE MONOMERIC
DIPHTHERIA TOXIN FUSION PROTEINS
AND THERAPEUTIC USES**

a continuation-in-part of application No. 16/083,848,
filed on Sep. 10, 2018, now Pat. No. 10,988,512, filed
as application No. PCT/US2017/021715 on Mar. 10,
2017.

(71) Applicants: **THE JOHNS HOPKINS
UNIVERSITY**, Baltimore, MD (US);
**TRUSTEES OF BOSTON
UNIVERSITY**, Boston, MA (US)

(60) Provisional application No. 62/306,281, filed on Mar.
10, 2016.

Publication Classification

(72) Inventors: **William R. Bishai**, Baltimore, MD
(US); **John R. Murphy**, Tilghman, MD
(US); **Laurene Cheung**, Baltimore, MD
(US); **Shashank Gupta**, Frederick, MD
(US); **Cynthia K. Bullen**, Baltimore,
MD (US)

(51) **Int. Cl.**
C07K 14/55 (2006.01)
C07K 14/34 (2006.01)
C12N 15/77 (2006.01)
(52) **U.S. Cl.**
CPC **C07K 14/55** (2013.01); **C07K 14/34**
(2013.01); **C12N 15/77** (2013.01); **A61K 38/00**
(2013.01)

(21) Appl. No.: **17/549,845**

(22) Filed: **Dec. 13, 2021**

Prior Publication Data

(15) Correction of US 2022/0204579 A1 Jun. 30, 2022
See (60) Related U.S. Application Data.

(65) US 2022/0204579 A1 Jun. 30, 2022

Related U.S. Application Data

(60) Division of application No. 16/611,243, filed on Nov.
6, 2019, now Pat. No. 11,203,626, filed as application
No. PCT/US2017/051020 on Sep. 11, 2017, which is

(57) **ABSTRACT**

The present invention is a DNA expression vector compris-
ing: a toxP; a mutant toxO that blocks Fe-mediated regula-
tion of gene expression; and a DNA sequence encoding a
protein, wherein the toxP and the mutant toxO regulate
expression of the DNA segment encoding the protein. It is
preferred that DNA expression vectors of the present inven-
tion include DNA sequences encoding a signal peptide so
that a protein expressed is attached to the signal peptide prior
to processing. Novel proteins are produced off of the DNA
expression vector of the present invention.

Specification includes a Sequence Listing.

toxO sequence and mutations

a) Mutant *toxO* in this invention report:

TTAGGATAGCTAAGTCCAT (altered bases shown in red)

b) Wild type *toxO*

TTAGGATAGCTTTACCTAA 19 bp imperfect palindrome around the large C

FIGs. 1A-1B *toxO* sequence and mutations

FIG. 1A

a) Mutant *toxO* in this invention report:

TTAGGATAGCTAAGTCCAT (altered bases shown in red)

FIG. 1B

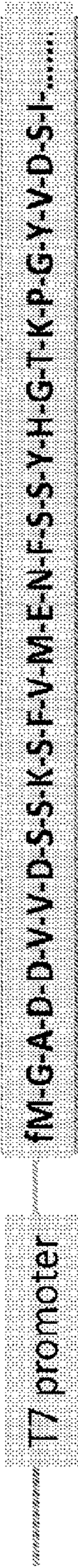
b) Wild type *toxO*

TTAGGATAGCTTTACCTAA 19 bp imperfect palindrome around the large C

FIGs. 2A-2B Addition of the *tox* promoter, mutant *tox* operator, and signal sequence in pKN2.6Z-LC127

FIG. 2A

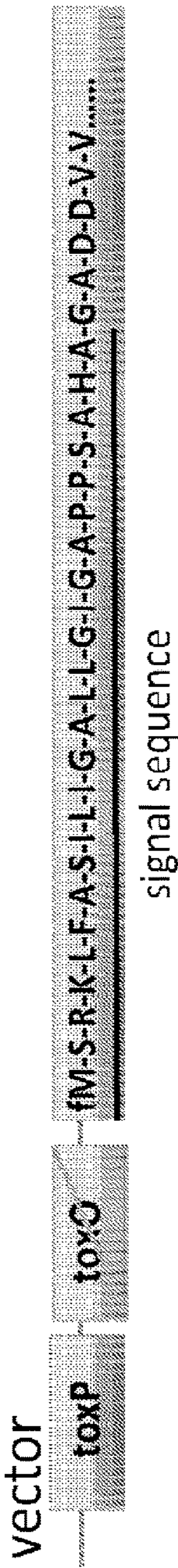
a) Classic denileukin difftitox (c-denileukin difftitox) expression vector



Classic Ontak® (c-Ontak®): Ontak® expressed in inclusion body form in the cytosol of recombinant *E. coli*. Must be denatured and refolded in the presence of Tween 20 to be active.

FIG. 2B

b) Secreted denileukin difftitox (s-denileukin difftitox)expression



Secreted Ontak® (s-Ontak®): expressed as a secreted protein (i) into the periplasm of recombinant *E. coli*, or (ii) into the spent culture medium of recombinant *C. diphtheriae*, C7s(-)*tox*-, as a fully folded biologically active protein. The red slash indicates that the *toxO* operator sequence has been mutated to prevent DtxR repressor binding. The mutant *toxO* enables constitutive, iron-independent expression of the downstream gene.

FIG. 3 c-denileukin diftotox-VLM shows similar activity to c-denileukin diftotox for killing IL-2R bearing cells

c-denileukin diftotox-VLM has a potency equivalent to c-denileukin diftotox for killing IL-2R bearing cells

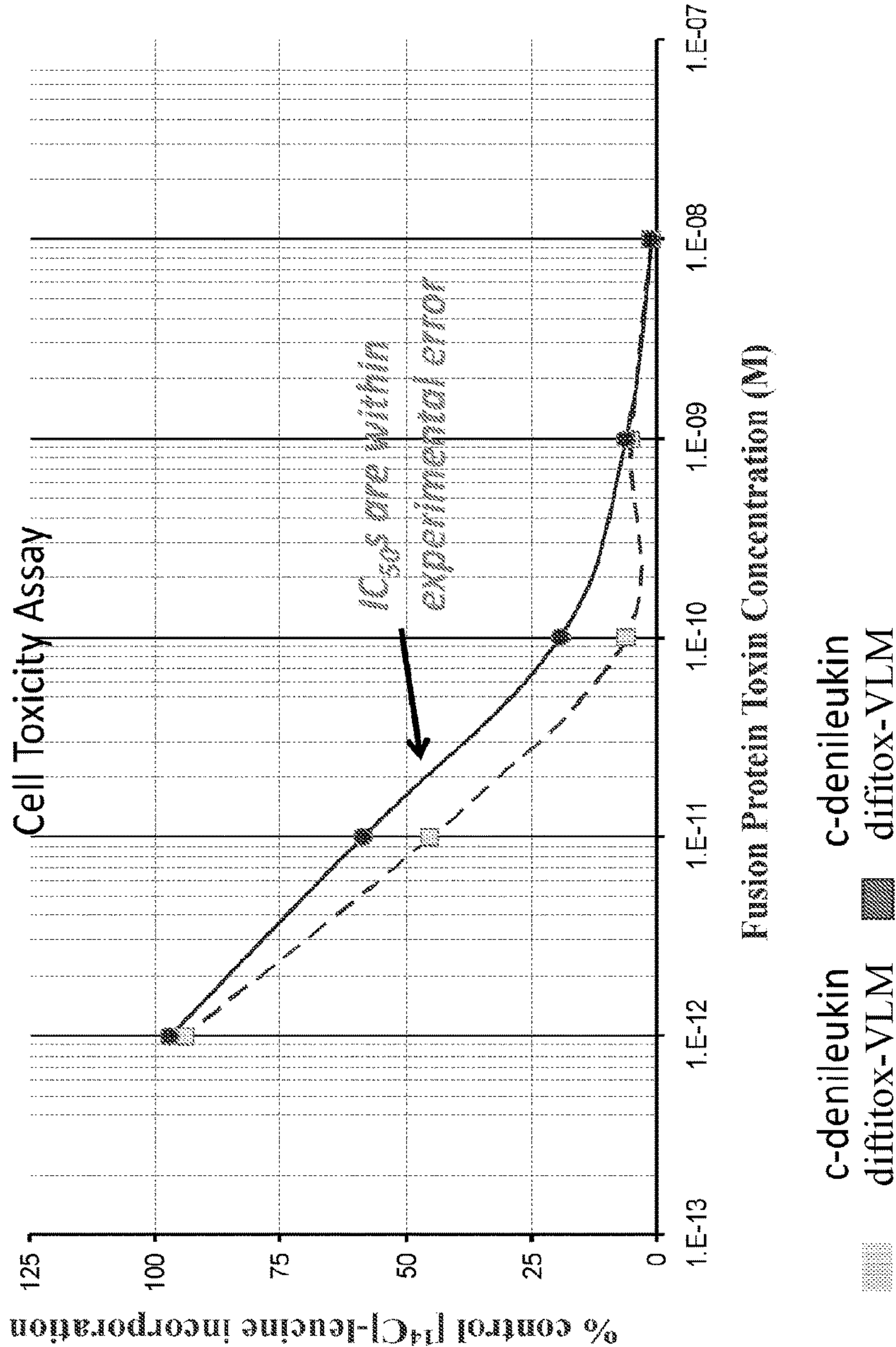


FIG. 4 c-denileukin diftitox-VLM : decreased
vascular leak

c-denileukin diftitox-VLM does not cause vascular leakage *in vitro*

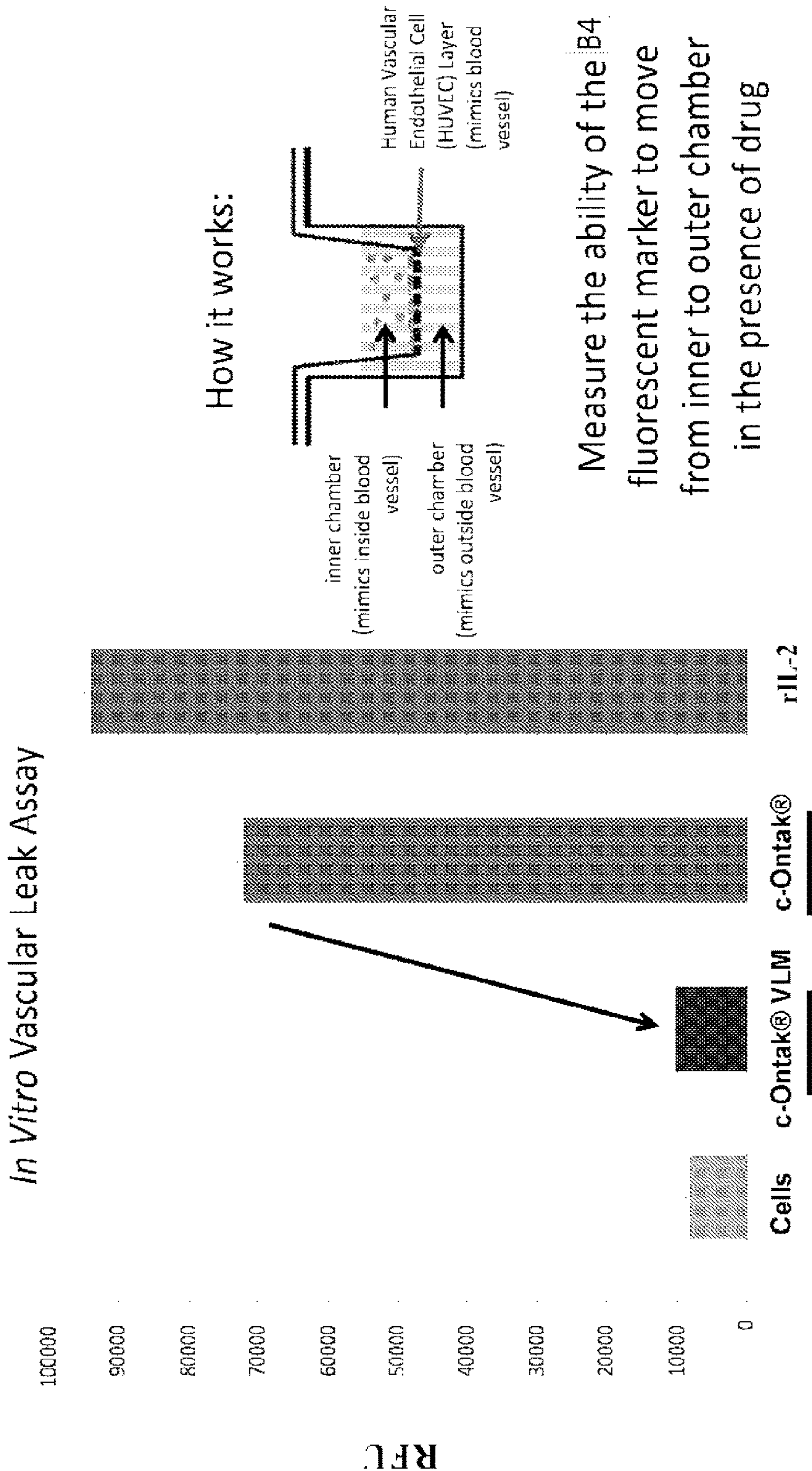


FIG. 5 c-denileukin diftitox-VLM: Safety and Tolerability

c-denileukin diftitox-VLM has significantly less acute toxicity *in vivo* than c-denileukin diftitox

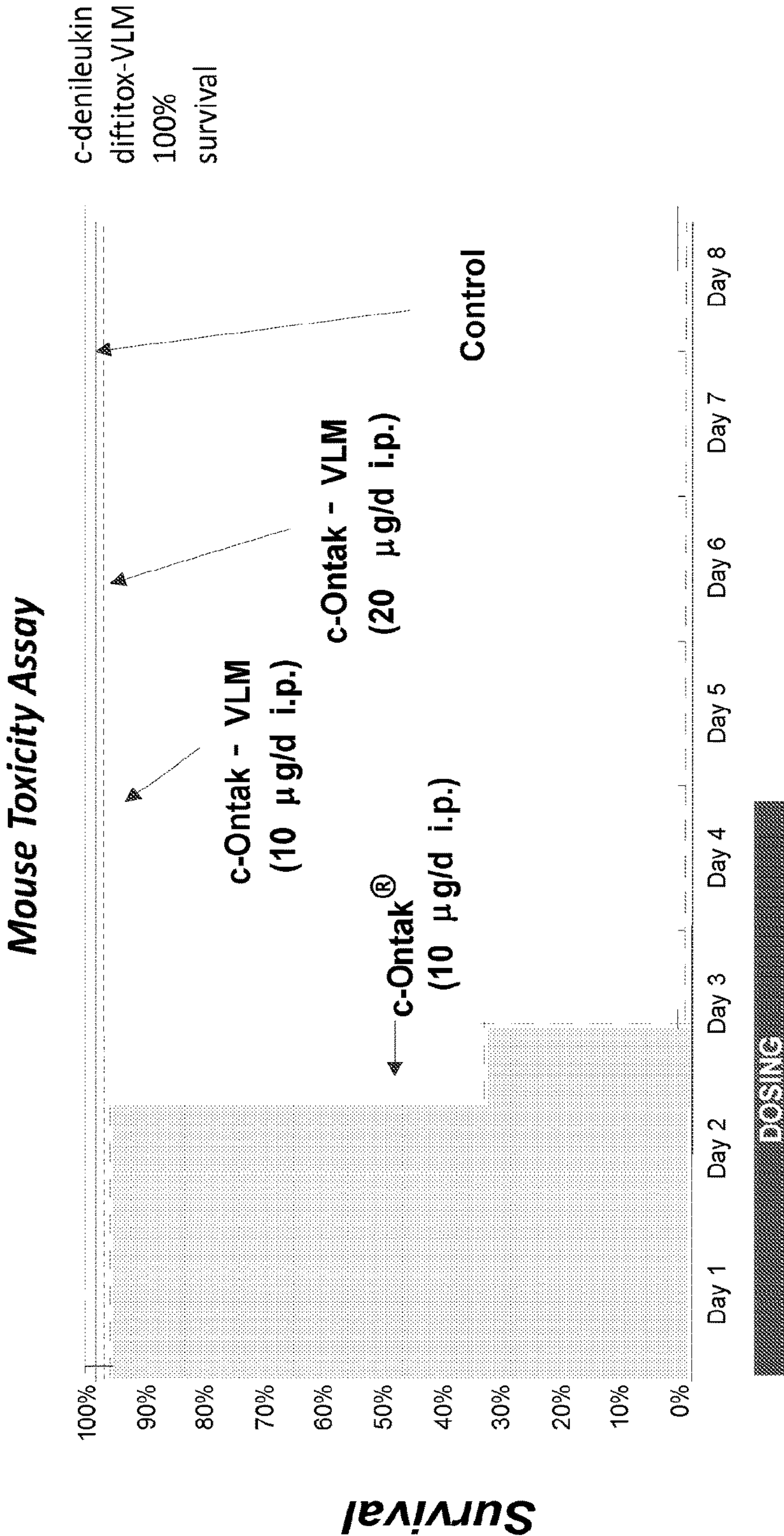


FIG. 6 *Diphtheria toxin-based fusion protein toxin platform technology*

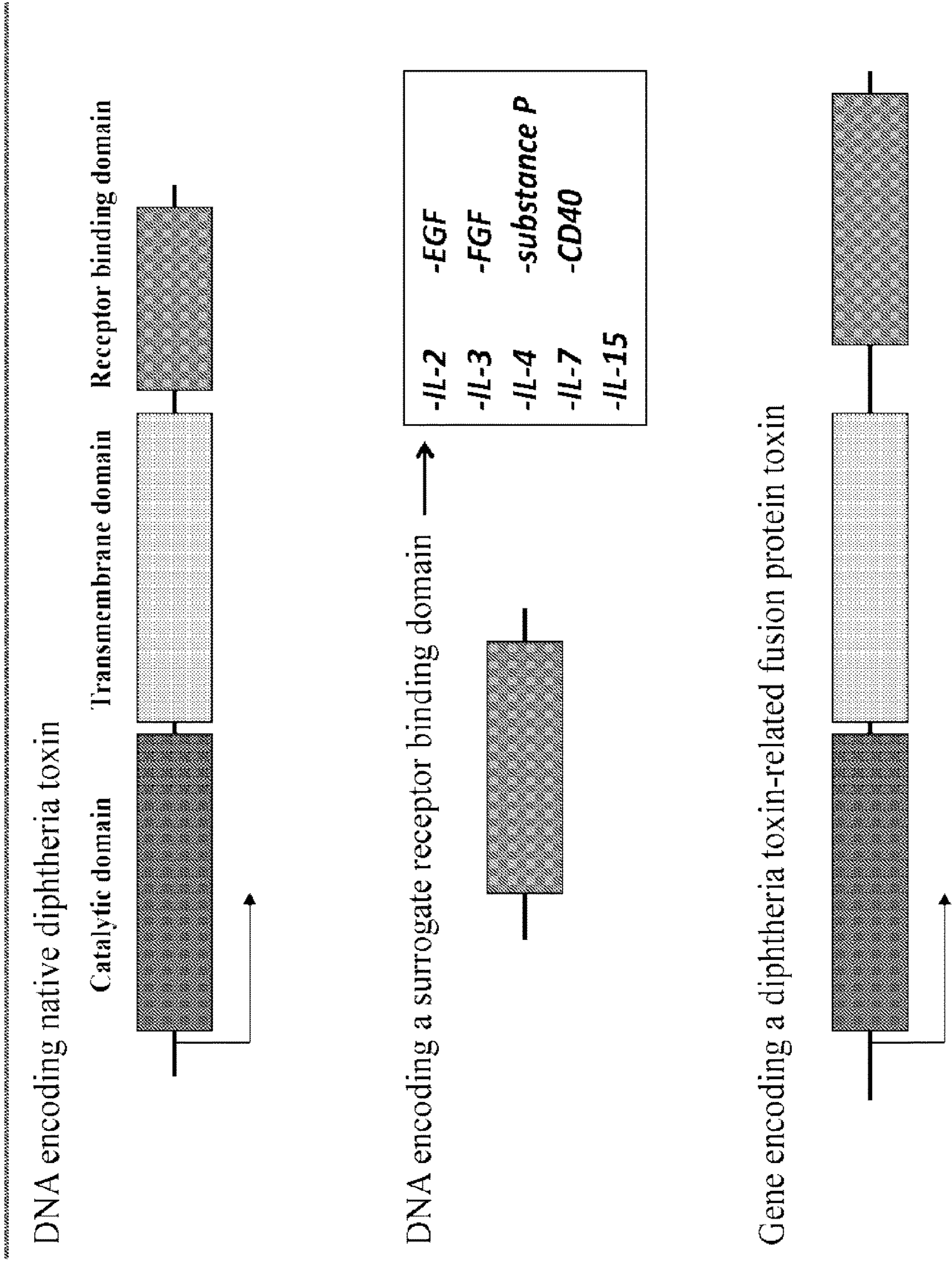
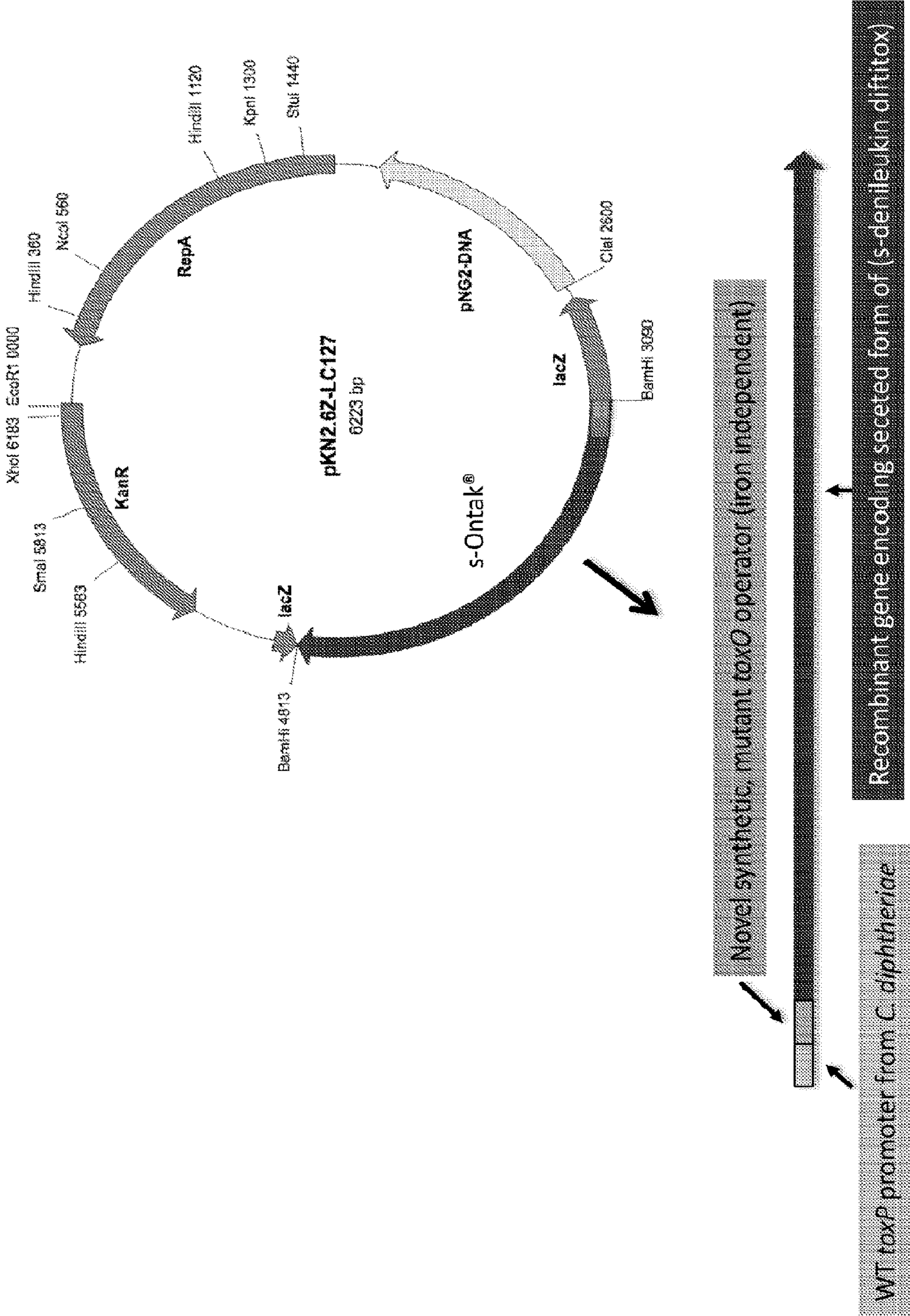
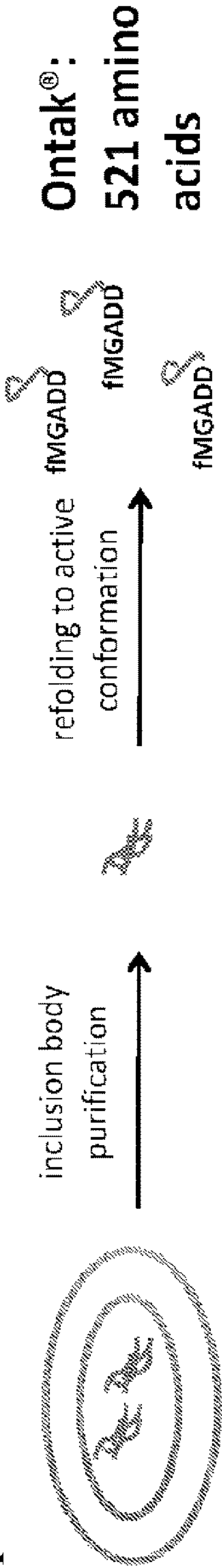


FIG. 7 Construct to express s-denileukin diftotox from *Corynebacterium diphtheriae* as a secreted protein into the culture medium in an iron-independent manner



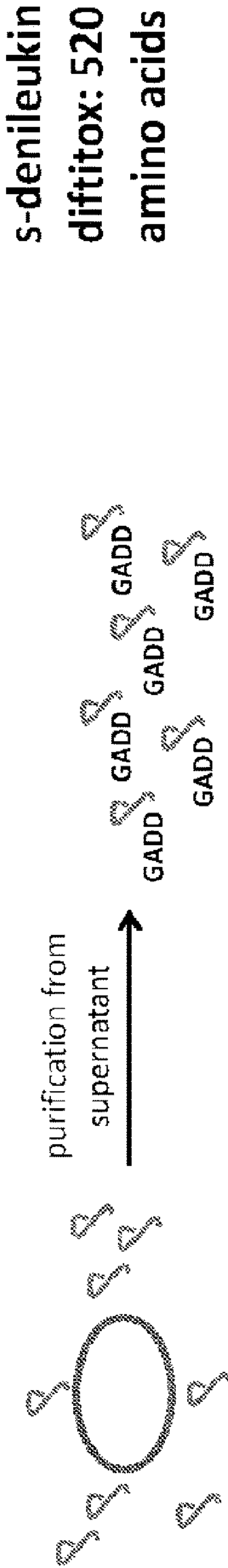
FIGs. 8A-8B Addressing the problems of inclusion body formation, denaturation, and refolding required for c-denileukin diftitox production

FIG. 8A



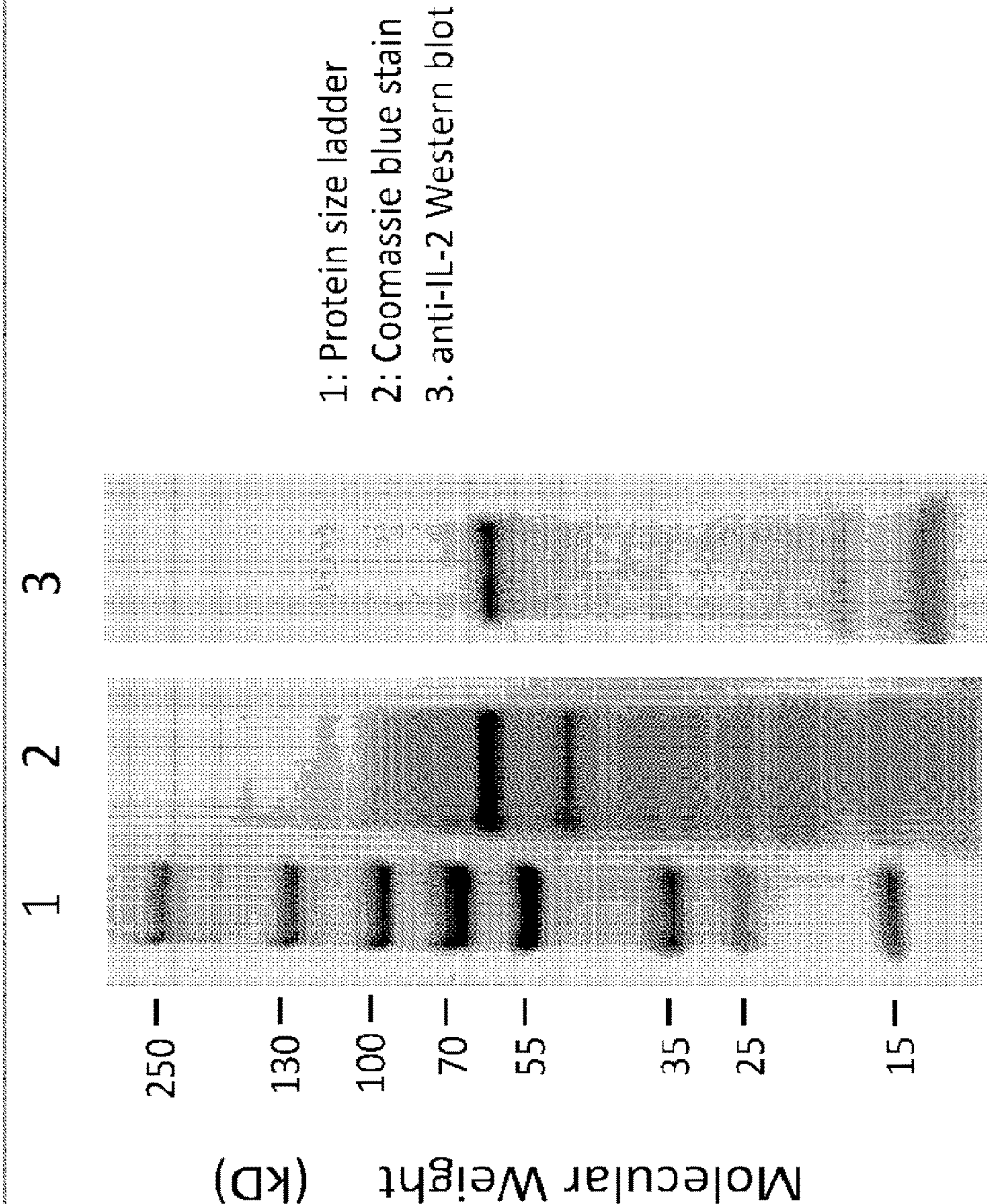
Ontak[®] is expressed in the *E. coli* cytoplasm and forms inclusion bodies; these must be denatured and refolded. Residual DNA, detergent, and protein aggregates are a limitation. Ontak[®] is 521 amino acids in length. Blue ovals indicate the gram-negative envelope of *E. coli*. fMGADD represent the first 5 amino acids of the N-terminus of Ontak[®].

FIG. 8B



s-denileukin diftitox is expressed by *C. diphtheriae* as a secreted protein into the cell free culture supernatant fluid in biologically active form. Mature s-denileukin diftitox is 520 amino acids in length. Red oval indicates the gram-positive envelope of *C. diphtheriae*. GADD represent the first 4 amino acids of the N-terminus of s-denileukin diftitox of the present invention.

FIG. 9 s-denilekin diftitox: expressed as a secreted protein into the culture medium of *Corynebacterium diphtheriae* strain C7(-) tox-



1

supernatant was concentrated 10x and 23 microliters of concentrate was loaded onto the gel

FIG. 10 Ontak® (DAB₃₈₉IL-2) is expected to deplete IL-2R-bearing (CD25+) T cells (T_{regs}) within the tuberculous granuloma. T_{regs} are immunosuppressive by their inhibition of T_{eff} cells

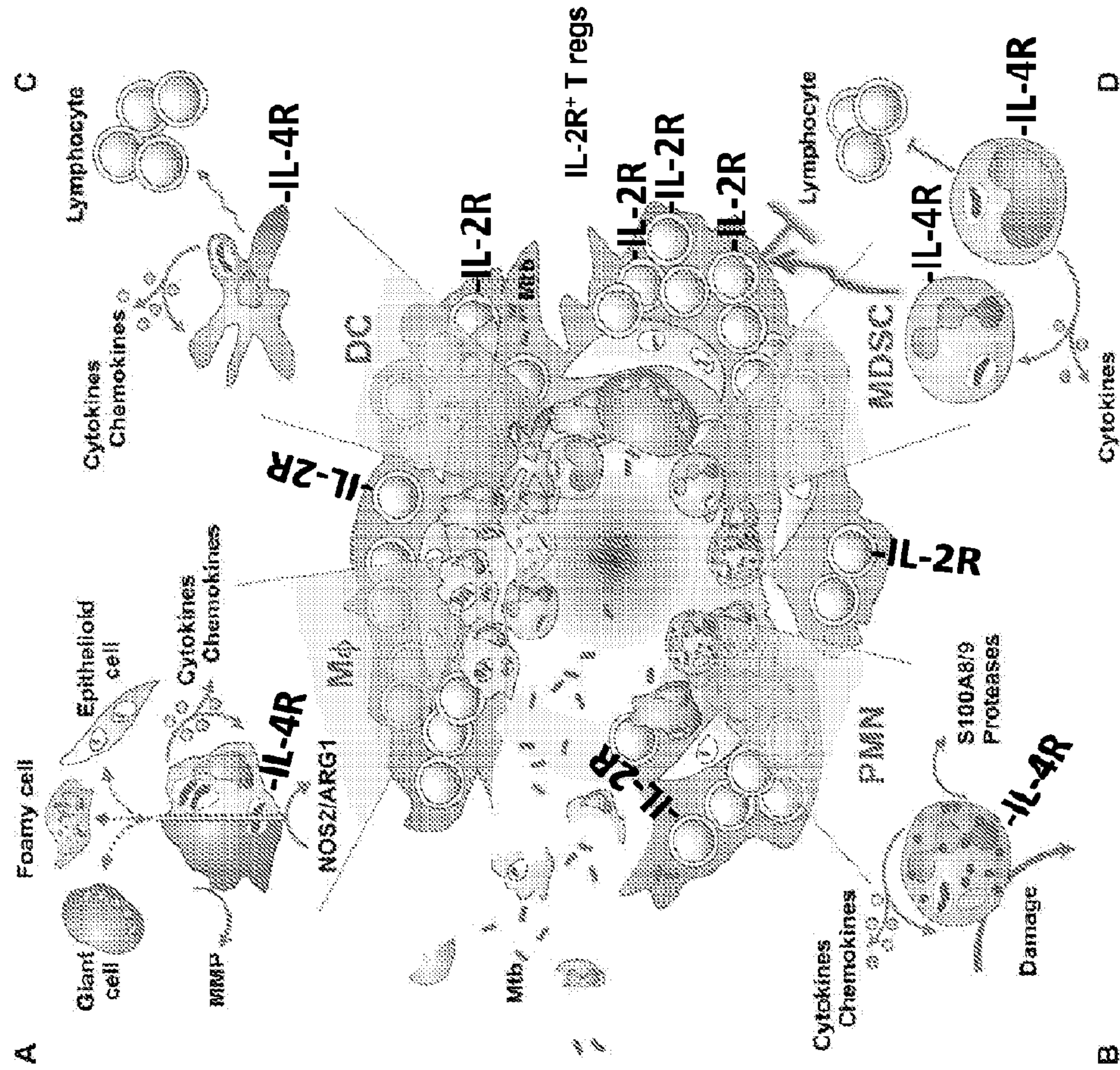


FIG 11. ONTAK for TB. Mouse TB Model. Experimental Scheme

Treatment	Week 0 No. of mice sacrificed for CFU counts	Week 2 No. of mice sacrificed for CFU counts	Week 3 No. of mice sacrificed for CFU counts
Grp 1. No treatment	5	5	5
Grp 2. Ont-2x IP	5	5	5
Grp 3. Ont-2x IV		5	5
Grp 4. Ont-1x IP		5	5
Grp 5. RHZ			5
Grp 6. RHZ + Ont-1x IP			5

Mice were infected with *M. tb.* strain H37Rv by aerosol infection giving an initial implantation of $\sim 2.8 \log_{10}$ CFU counts in lungs on day 0. The groups of mice were treated with 750 ng of c-denileukin diftiox intraperitoneally (IP) or intravenously (IV) as one treatment cycle (1x, dosed at week 2 post-infection) or two treatment cycles (2x, dosed at ~day 3 pre-infection and week 2 post-infection). A treatment cycle of c-denileukin diftiox is defined as 35 $\mu\text{g/kg}$ (750 ng for a typical mouse) given two times, two days apart. RHZ daily treatment by oral gavage was started at week 2. R – rifampin 10 mg/kg, I – isoniazid 10 mg/kg, Z – pyrazinamide 150 mg/kg.

FIG. 12 Ontak to treat TB. Mouse TB Model
Lung CFUs during treatment with different ONTAK regimens

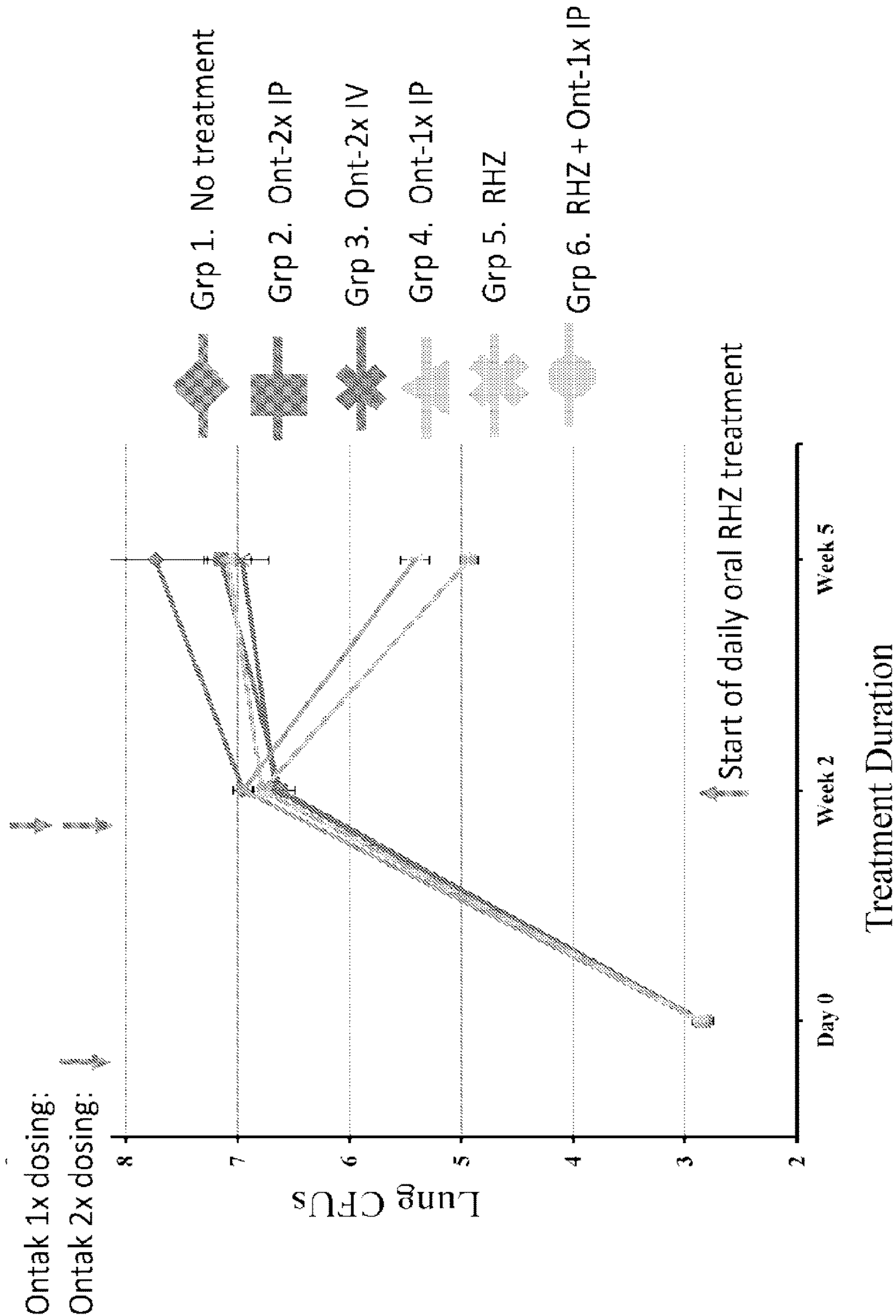


FIG 13. Ontak to treat TB. Mouse TB Model
Lung CFUs during treatment with different c-denileukin diftiox regimens

Data from Figure 12 with just Group 1 (no treatment) and Group 4 (c-denileukin diftiox 1x or one treatment cycle as monotherapy IP at week 2 post-infection)

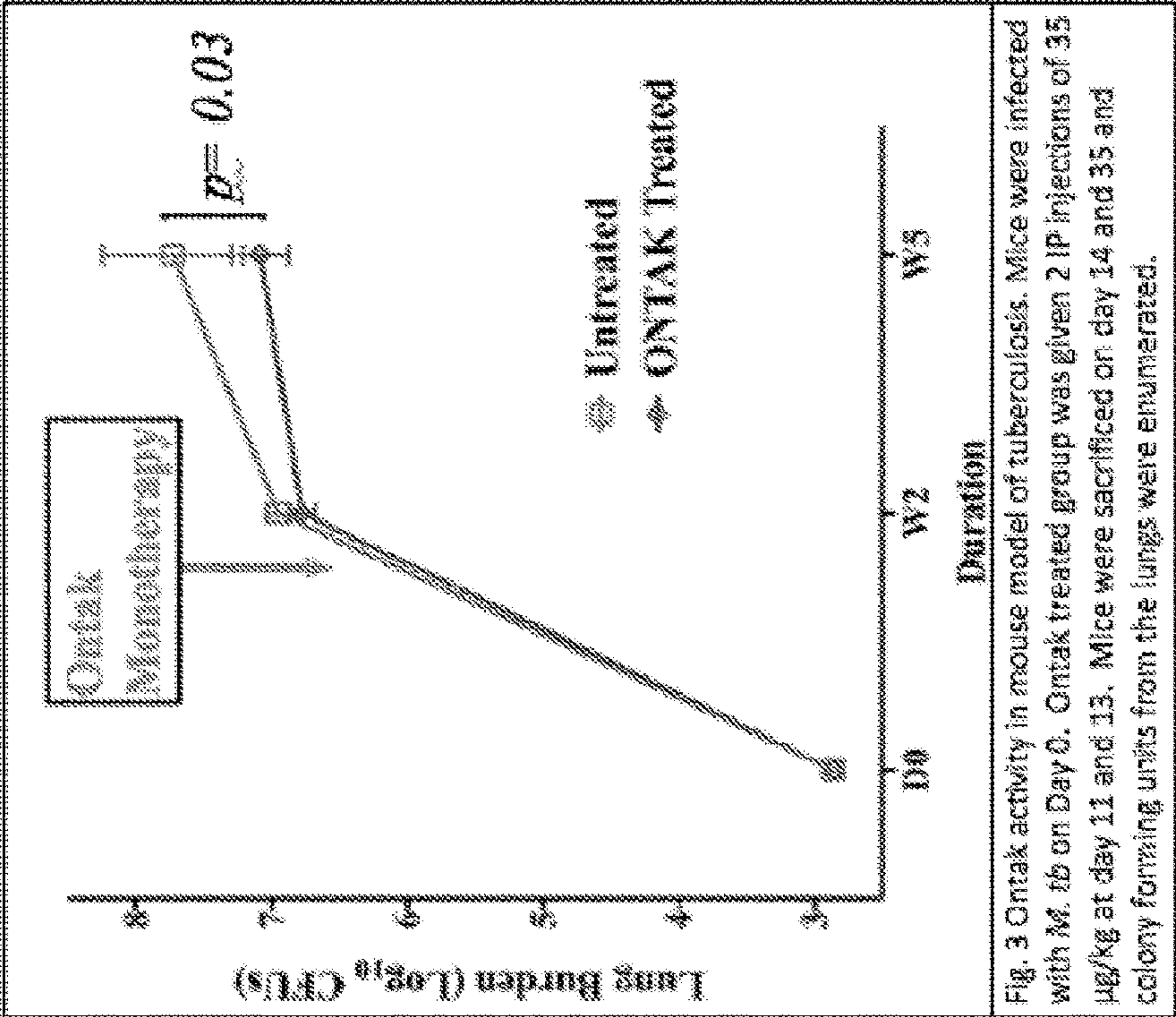
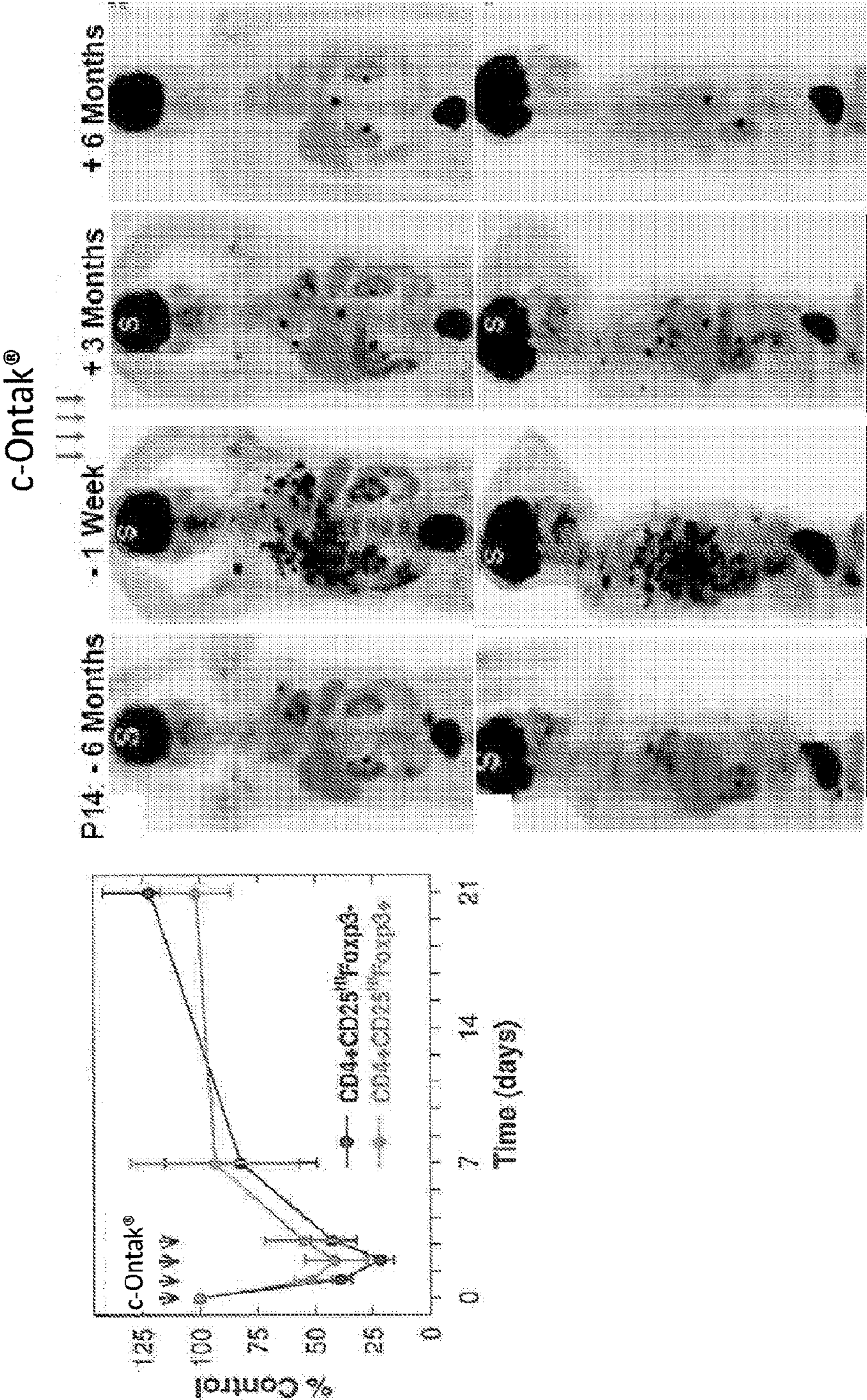


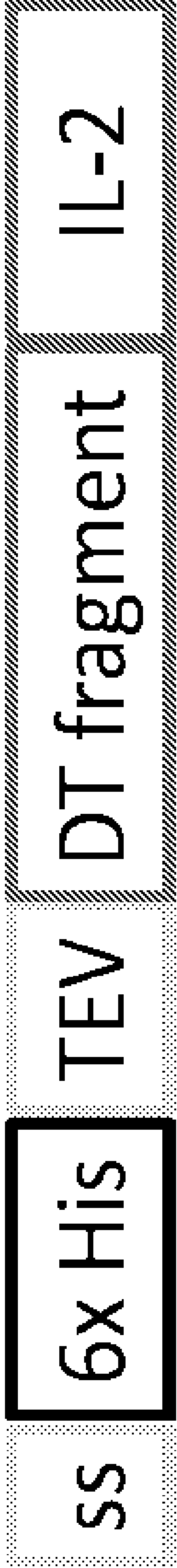
FIG. 14 c-denileukin diffitox as an immunotherapeutic agent: Malignant Melanoma



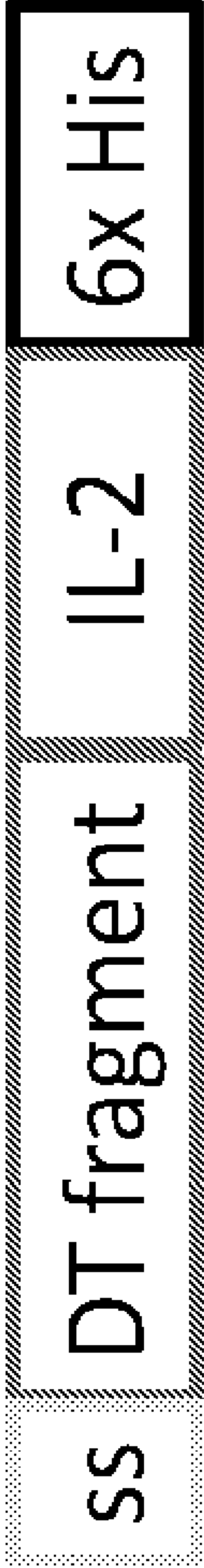
1 cycle = 12 μ g/kg QD x 4; rest 21 days

FIG. 15 Ontak production: His tag

N-terminal: His₆-TEV-
VLM s-Ontak



C-terminal: His₆-
VLM s-Ontak



C terminus TEV-His₉-
VLM s-Ontak

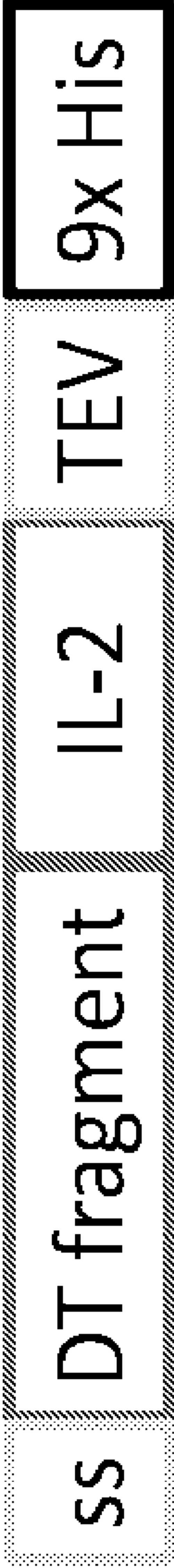


FIG 16. *VLM s-Ontak:*
Purification to > 97% with C-terminal His6-tagged VLM s-Ontak

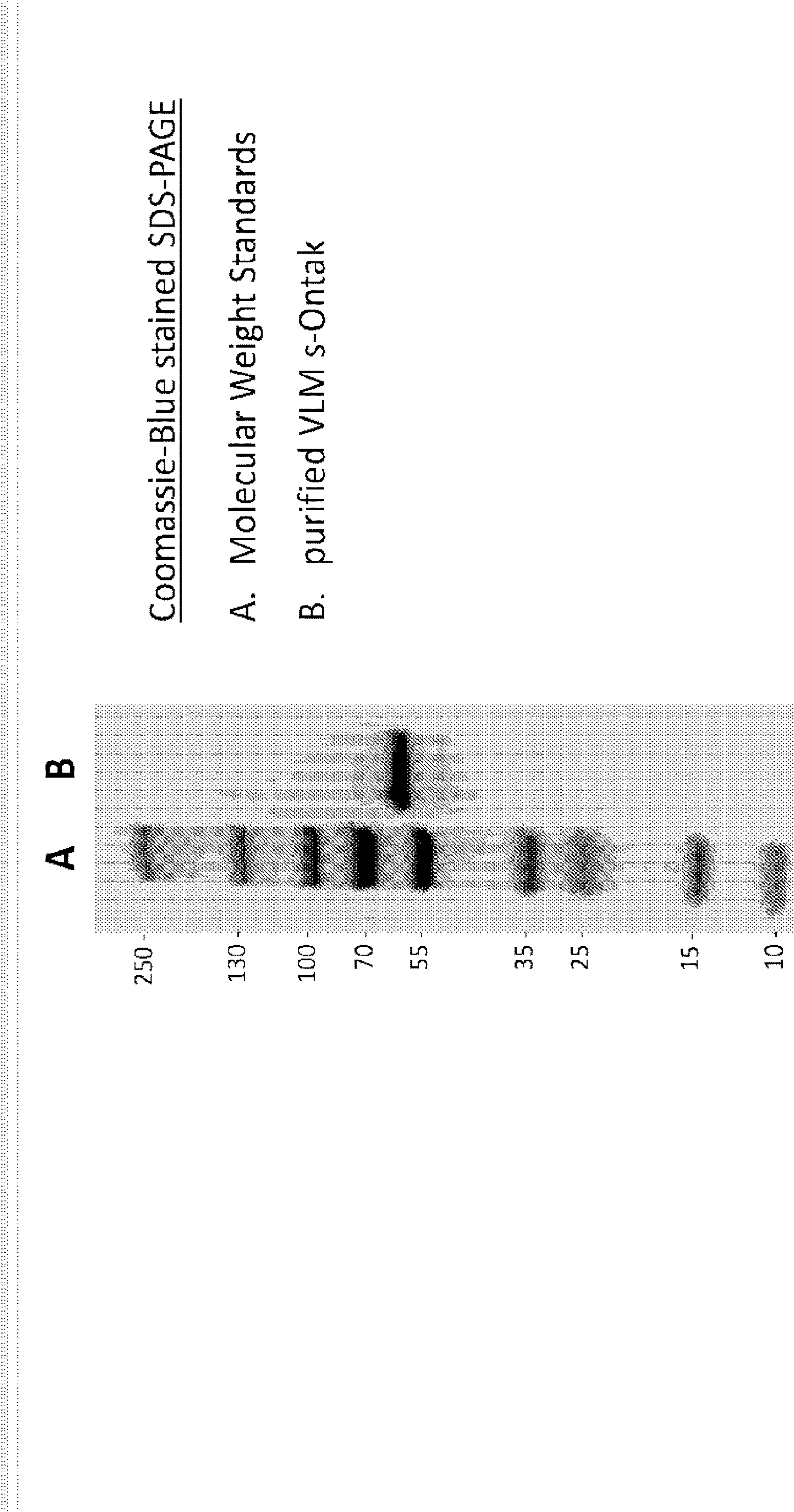
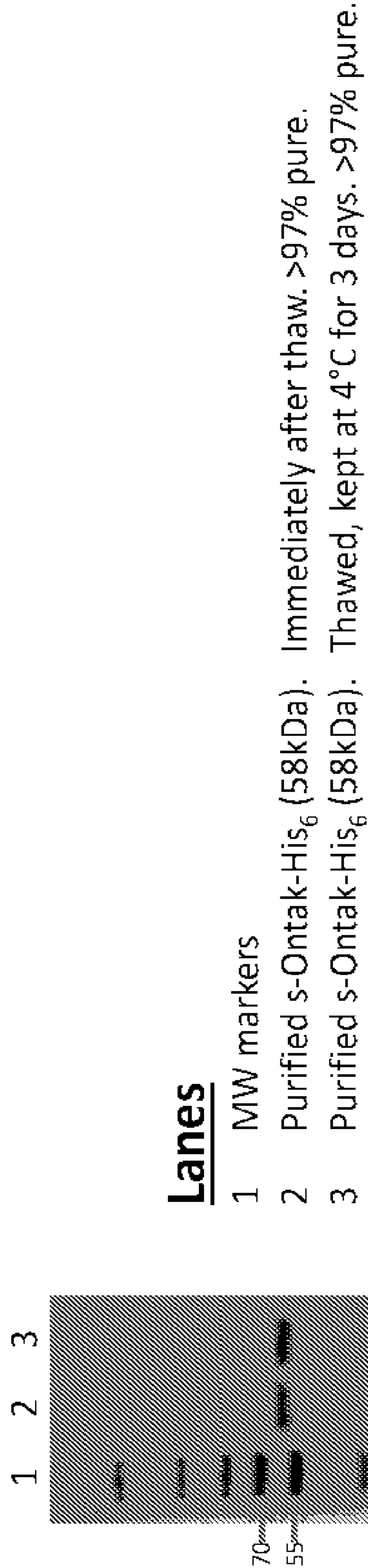


FIG. 17

Purified s-Ontak-His₆ (*C. diphtheriae*-derived)



SDS-PAGE with Coomassie-blue staining

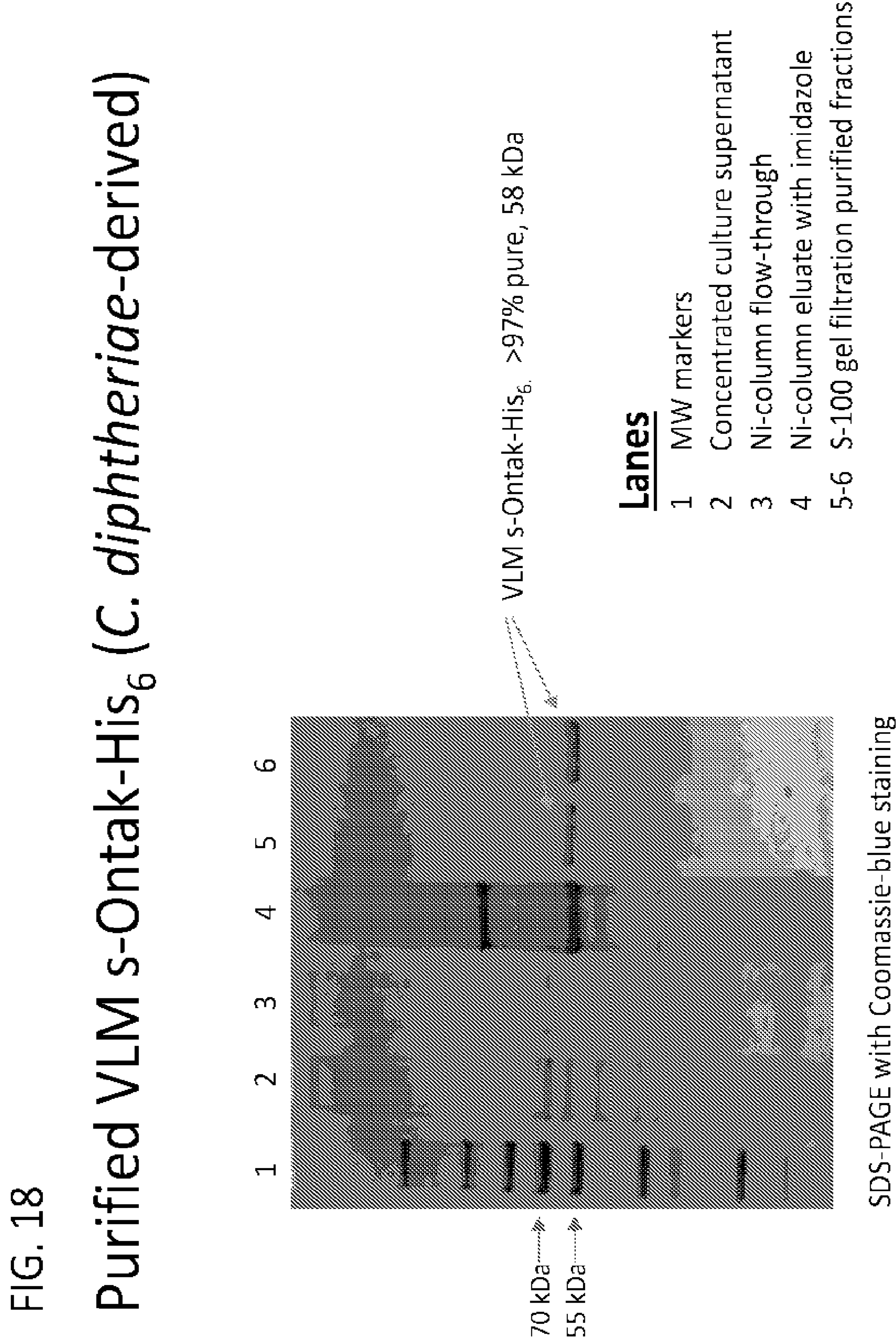
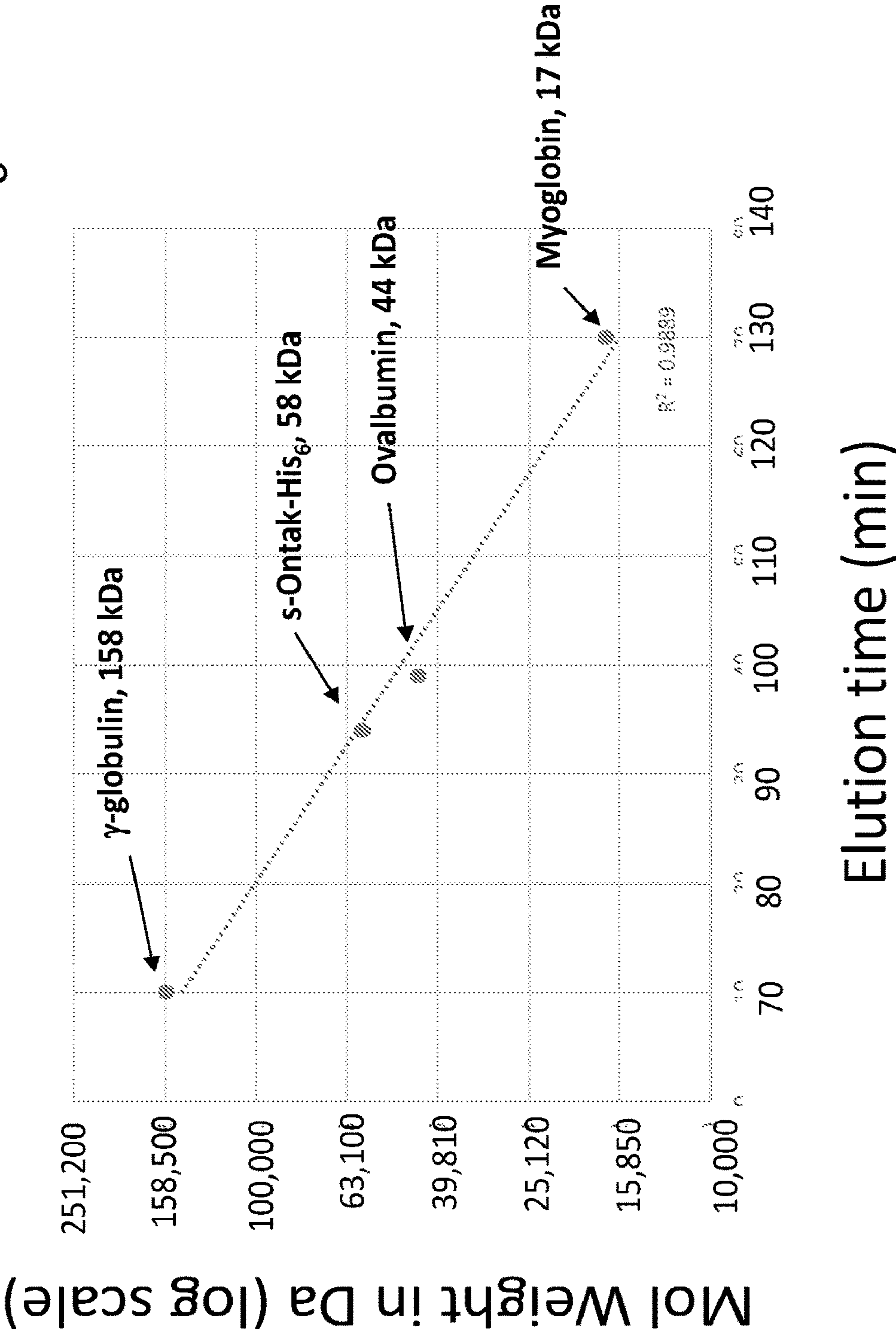


FIG. 19

S-100 Gel Filtration of s-Ontak-His₆



METHODS OF PRODUCING AGGREGATE-FREE MONOMERIC DIPHThERIA TOXIN FUSION PROTEINS AND THERAPEUTIC USES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a Divisional Application of U.S. application Ser. No. 16/611,243 filed Nov. 6, 2019, now pending; which is a 371 National Stage Application of International Application No. PCT/US2017/051020 filed Sep. 11, 2017, now expired and claims priority to International Application No. PCT/US2017/021715 filed Mar. 10, 2017; U.S. application Ser. No. 16/611,243 is also a Continuation-in-Part Application of U.S. application Ser. No. 16/083,848 filed Sep. 10, 2018, now issued as U.S. Pat. No. 10,988,512; which is a 371 National Stage Application of International Application No. PCT/US2017/021715 filed Mar. 10, 2017, which claims the benefit under 35 USC § 119(e) to U.S. Application Ser. No. 62/306,281 filed Mar. 10, 2016, now expired. The disclosure of each of the prior applications is considered part of and is incorporated by reference in the disclosure of this application.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under Grant Nos. A1037856, A1036973, A1097138 and UC7A1095321-01 awarded by the National Institutes of Health. The government has certain rights in the invention.

INCORPORATION BY REFERENCE OF SEQUENCE LISTING

[0003] The material in the accompanying sequence listing is hereby incorporated by reference into this application. The accompanying sequence listing text file, named JHU4270-3_Seq.Listing.txt, was created on Dec. 13, 2021 and is 140 kb in size. The file can be accessed using Microsoft Word on a computer that uses Windows OS.

BACKGROUND OF THE INVENTION

[0004] Ontak® (denileukin diftitox), is a 521 amino acid, recombinant, DNA-derived cytotoxic protein composed of the sequences for diphtheria toxin fragments A and a portion of fragment B (Meti-His₃₈₈) and the sequences for human interleukin-2 (IL-2; Alai-Thri33). It is currently produced in an *E. coli* expression system and has a molecular weight of 58 kD. Neomycin is used in the fermentation process but is undetectable in the final product. Ontak®, which is supplied in single use vials as a sterile, frozen solution intended for intravenous (IV) administration, was approved by the FDA in 1999 for the treatment of cutaneous T cell lymphoma (CTCL). The FDA placed Ontak® on clinical hold in June 2011 because of concerns regarding the presence of protein aggregates of heterogeneous molecular weight, excess residual DNA, and excess residual detergent in the final formulation. The production of Ontak® was achieved by expressing the recombinant protein in the *E. coli* cytoplasm, and this expression system resulted in the recombinant protein forming large insoluble aggregates or so-called inclusion bodies comprised of the Ontak® polypeptide. In the current process of production, which includes denaturation and refolding of the inclusion body forms, protein

aggregates of heterogeneous molecular weight were still present in the final formulation. The presence of these aggregates in the purified form is a consequence of using *E. coli*-derived cytoplasmic inclusion bodies as the source of the polypeptide and because of the intrinsic hydrophobic nature of the toxin's transmembrane domain even in the presence of Tween 20. Ontak® produced using this method will hereafter be referred to as classic-Ontak® or c-Ontak®.

[0005] In addition, like all of the bacterial and plant toxins, c-Ontak® carries amino acid motifs that induce vascular leak syndrome (VLS). Approximately 30% of patients treated with c-Ontak® develop VLS symptoms ranging from peripheral edema with rapid weight gain to hypoalbuminemia to pulmonary edema. What is needed are 1) a process enabling the production of Ontak-like proteins at high yields and purity, eliminating aggregates in the final commercial product, and 2) modified Ontak-like proteins with minimal VLS side-effects to provide safer drugs to patients.

SUMMARY OF THE INVENTION

[0006] One embodiment of the present invention is a DNA expression vector comprising: a toxP; a mutant toxO that blocks Fe-mediated regulation of gene expression; and a DNA sequence encoding a protein, wherein the toxP and the mutant toxO regulate expression of the DNA segment encoding the protein. It is preferred that DNA expression vectors of the present invention include DNA sequences encoding a signal peptide so that a protein expressed off a DNA expression vector is attached to the signal peptide that is typically cleaved off to form a mature protein. The preferred mutant toxO is SEQ ID NO: 1 and the preferred signal peptide is SEQ ID NO: 5. The DNA expression vectors of the present invention may be used to produce many kinds of proteins such as CRM 197 and CRM 107, or a combination thereof. CRM protein sequences are illustrated in SEQ ID NOs: 18-21. It is preferred that the DNA expression vectors of the present invention encode a diphtheria toxin, or functional part thereof, attached to a receptor binding protein, or a functional part thereof to form a diphtheria toxin receptor fusion protein. The receptor binding protein portion of such fusion proteins may be selected from the group comprising IL-2, IL-3, IL-4, IL-6, IL-7, IL-15, EGF, FGF, substance P, CD4, αMSH, GRP, TT fragment C, GCSF, heregulin β1, a functional part thereof, or a combination thereof. Examples of diphtheria toxin fusion proteins include the proteins illustrated in any one of SEQ ID NOs: 11-15, 30, 38-40, 42-43, 45-46, and 58, and proteins encoded by a nucleic acid of any one of SEQ ID NOs: 31, 41, 44, and 59.

[0007] Another embodiment of the present invention is a DNA expression vector comprising: a toxP; a mutant toxO that blocks Fe-mediated regulation of gene expression; a DNA sequence encoding a protein comprising a signal sequence, a diphtheria toxin, or a functional part thereof, that is free of a diphtheria receptor binding domain or has a non-functional diphtheria toxin receptor binding domain, and a target receptor binding domain selected from the group comprising IL-2, IL-3, IL-4, IL-6, IL-7, IL-15, EGF, FGF, substance P, CD4, αMSH, GRP, TT fragment C, GCSF, heregulin β1, a functional part thereof, or a combination thereof, wherein the toxP and the mutant toxO regulate expression of the DNA sequence encoding the protein. Typically, a bacteria transformed with a DNA expression

vector of the present invention produces a diphtheria toxin receptor binding fusion protein attached to a signal peptide that is directed to a periplasm, a culture medium, or both locations by the signal peptide. If the bacteria is *E. coli* then the signal peptide typically directs the diphtheria toxin receptor binding fusion protein to the periplasm. If the bacteria is *Corynebacterium diphtheria* then signal peptide directs the diphtheria toxin receptor binding fusion protein to the culture medium. It is preferred that a DNA expression vector of the present invention comprises SEQ ID NO: 3 and may comprise a DNA encoding a cleavable protein tag wherein the cleavable protein tag is attached to the diphtheria toxin receptor binding fusion protein. Example of diphtheria toxin receptor binding fusion proteins produced from the DNA expression vectors of the present invention include any one of SEQ ID NOs: 11-15, 30, 38-40, 42-43, 45-46, and 58, and proteins encoded by a nucleic acid of any one of SEQ ID NOs: 31, 41, 44, and 59.

[0008] Another embodiment of the present invention includes a method for producing aggregate-free monomeric diphtheria toxin fusion proteins comprising the following steps: transforming bacteria with a DNA expression vectors of the present invention; forming a transformant; incubating the transformant in a culture medium to allow expression of a protein that is secreted into the culture medium; and purifying the protein from the culture medium. The preferred bacteria used in this method is *Corynebacterium diphtheria*.

[0009] Another embodiment of the present invention includes a method for producing aggregate-free monomeric diphtheria toxin fusion proteins comprising the following steps: 1) transforming *Corynebacterium diphtheria* strain with a DNA vector comprising: a toxP; a mutant toxO that blocks Fe-mediated regulation of gene expression; a DNA sequence encoding a protein comprising: signal peptide; a diphtheria toxin, or a functional part thereof, that is free of a diphtheria receptor binding domain or has a non-functional diphtheria toxin receptor binding domain; and a target receptor binding domain selected from the group comprising IL-2, IL-3, IL-4, IL-6, IL-7, IL-15, EGF, FGF, substance P, CD4, α MSH, GRP, TT fragment C, GCSF, heregulin β 1, TNF α , TGF β , a functional part thereof, or a combination thereof, wherein the toxP and the mutant toxO regulate expression of the DNA sequence encoding the protein; 2) forming a transformant; 3) incubating the transformant in a culture medium to allow expression of the protein and that is secreted into the culture medium; and 4) purifying the diphtheria toxin fusion protein from the culture medium. Examples of diphtheria toxin receptor fusion proteins produced by methods of the present invention include any one of SEQ ID NOs: 11-15, 30, 38-40, 42-43, 45-46, and 58, and proteins encoded by a nucleic acid of any one of SEQ ID NOs: 31, 41, 44, and 59. The preferred *Corynebacterium diphtheria* strain used in the methods of the present invention is *Corynebacterium* C7 beta (-), tox (-).

[0010] Another embodiment of the present invention includes a method of treating a patient with tuberculosis comprising the following steps: preparing a diphtheria toxin fusion protein as provided in this application; administering the diphtheria toxin fusion protein to a patient with tuberculosis.

[0011] Another embodiment of the present invention includes a DNA expression vector comprising a mutant toxO promoter.

[0012] Another embodiment of the present invention includes a *Corynebacterium diphtheria* strain containing a DNA expression vector of the present invention.

[0013] Another embodiment of the present invention is method of making a protein comprising the following steps: providing a DNA expression vector comprising a toxP, a mutant /OA) that blocks Fe-mediated regulation of gene expression, a signal sequence, and a DNA sequence encoding a protein; transforming a bacteria strain with the DNA vector to form a transformant; incubating the transformant in a culture medium for a period of time to allow expression of a protein that is secreted into the culture medium; and purifying the protein from the culture medium.

[0014] Another embodiment of the present invention is a fusion protein selected from any one of SEQ ID NOs: 11-15, 30, 38-40, 42-43, 45-46, and 58, or encoded by a nucleic acid of any one of SEQ ID NOs: 31, 41, 44, and 59.

[0015] Another embodiment of the present invention is a pharmaceutical composition comprising a fusion protein described above.

[0016] Another embodiment of the present invention is a pharmaceutical composition comprising a fusion protein describe above, and at least one or more other chemotherapy agents. Examples of chemotherapy agents include isoniazid, rifampin, rifabutin, rifapentine, pyrazinamide, ethambutol, streptomycin, amikacin, kanamycin, ethionamide, protionamide, terizidone, thiacetazone, cycloserine, capreomycin, para-amino salicylic acid (PAS), viomycin, ofloxacin, ciprofloxacin, levofloxacin, moxifloxacin, bedaquiline, or delamanid, linezolid, tedezolid, amoxicillin-clavulanic acid, meropenem, imipenem, clarithromycin or clofazimine.

[0017] A pharmaceutical composition of comprising a fusion protein described above, and at least one or more other antimicrobial agents. Examples of antimicrobial agents include isoniazid, rifampin, rifabutin, rifapentine, pyrazinamide, ethambutol, streptomycin, amikacin, kanamycin, ethionamide, protionamide, terizidone, thiacetazone, cycloserine, capreomycin, para-amino salicylic acid (PAS), viomycin, ofloxacin, ciprofloxacin, levofloxacin, moxifloxacin, bedaquiline, or delamanid, linezolid, tedezolid, amoxicillin-clavulanic acid, meropenem, imipenem, clarithromycin, or clofazimine.

[0018] Another embodiment of the present invention is a method of treating or preventing cancer in a subject comprising administering to the subject an effective amount of a pharmaceutical composition comprising a fusion protein selected from any one of SEQ ID NOs: 11-15, 30, 38-40, 42-43, 45-46, and 58, or encoded by a nucleic acid selected from any one of SEQ ID NOs: 31, 41, 44, and 59.

[0019] Another embodiment of the present invention is a method of treating or preventing tuberculosis in a subject comprising administering to the subject an effective amount of a pharmaceutical composition comprising a fusion protein selected from any one of SEQ ID NOs: 11-15, 30, 38-40, 42-43, 45-46, and 58, or encoded by a nucleic acid selected from any one of SEQ ID NOs: 31, 41, 44, and 59.

[0020] Another embodiment of the present invention is a prokaryotic cell line comprising a DNA expression vector of the present invention.

[0021] Another embodiment of the present invention is kit comprising the DNA expression vector of the present invention.

[0022] Another embodiment of the present invention is a toxP comprising SEQ ID NO: 2.

[0023] Another embodiment of the present invention is a protein of any one of SEQ ID NOs: 11-15, 30, 38-40, 42-43, 45-46, and 58, or a protein encoded by a nucleic acid selected from any one or SEQ ID NOs: 31, 41, 44, and 59.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] FIGS. 1a-1b illustrate: 1a) a mutant toxO of the present invention (SEQ ID NO: 1) and 1b) a wild type toxO (SEQ ID NO: 25) and a DtxR consensus binding sequence.

[0025] FIGS. 2a-2b illustrate: 2a) the classic denileukin diftotox (c-denileukin diftotox) expression vector used to manufacture Ontak® and 2b) the secreted denileukin diftotox (s-denileukin diftotox) expression vector including the tox promoter (toxP), and mutant toxO of the present invention. FIG. 2a discloses SEQ ID NO: 26 and FIG. 2b discloses SEQ ID NO: 27.

[0026] FIG. 3 illustrates a vascular leak mutant (VLM) called c-denileukin diftotox-VLM has equivalent potency to c-denileukin diftotox in killing IL2-receptor-bearing cells.

[0027] FIG. 4 illustrates c-denileukin diftotox-VLM does not cause vascular leak in vitro.

[0028] FIG. 5 illustrates that c-denileukin diftotox-VLM has significantly less acute toxicity in vivo than c-Ontak® using a mouse survival model.

[0029] FIG. 6 illustrates a diphtheria toxin-based fusion protein toxin platform technology of the present invention.

[0030] FIG. 7 illustrates plasmid pKN2.6Z-LC127 with the tox promoter (toxP of SEQ ID NO: 2) and a mutant tox operator (toxO) (DNA SEQ ID NO: 1), a signal peptide (DNA SEQ ID NO: 4) attached to c-denileukin diftotox DNA sequences (DNA SEQ ID NO: 6).

[0031] FIGS. 8a-8b illustrate: 8a) the problems of the conventional process of manufacturing Ontak® as cytoplasmic inclusion bodies in *E. coli* and 8b) illustrates easy and clean manufacturing process of producing a secreted-denileukin diftotox having one less amino acid than the Ontak® protein. FIG. 8a discloses “fMGADD” as SEQ ID NO: 28 and FIG. 8b discloses “GADD” as SEQ ID NO: 29.

[0032] FIG. 9 illustrates an immunoblot of s-denileukin diftotox prepared by the process of the present invention where s-denileukin diftotox is expressed within a *Corynebacterium diphtheria* strain C7 beta (-), tox (-) and is secreted into the culture medium.

[0033] FIG. 10 illustrates how a denileukin diftotox of the present invention, is expected to deplete IL-2R (CD25+) bearing T cells (T_{regs}) within a tuberculosis granuloma. T_{regs} are immunosuppressive by their inhibition of T_{eff} cells.

[0034] FIG. 11 illustrates diphtheria fusion proteins used in the in vivo treatment of subjects (mice) with *M. tuberculosis*.

[0035] FIG. 12 illustrates the results of treating subjects (mice) infected with *M. tuberculosis* with diphtheria toxin-based fusion proteins.

[0036] FIG. 13 illustrates a diphtheria toxin-based fusion protein regimen for treating subjects (mice) infected with *M. tuberculosis*.

[0037] FIG. 14 illustrates the use of a diphtheria toxin-based fusion protein to treat subjects (humans) with malignant melanoma.

[0038] FIG. 15 illustrates the three constructs for rapid production of VLM s-Ontak and related proteins using His (histidine tags) (“His₆/6x His” and “His₉/9x His” disclosed as SEQ ID NOs: 23 and 48, respectively).

[0039] FIG. 16 illustrates purified VLM s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) at greater than 97% purity produced using the C-terminal His₆ VLM s-Ontak construct (“His₆” disclosed as SEQ ID NO: 23). Specifically, a recombinant *C. diphtherias* harboring a gene construct encoding VLM s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was grown to an optical density (OD) of ~12. The culture supernatant was harvested, concentrated by tangential flow filtration using a 30 kDa molecular weight cut-off membrane, and diafiltered for buffer exchange using tangential flow filtration as above. The protein mixture was partially purified by Ni-affinity chromatography and then purified to greater than 97% by gel permeation chromatography using S-100 resin. The resulting VLM s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was >97% pure.

[0040] FIG. 17 illustrates purified s-Ontak at greater than 97% purity produced using the C-terminal His₆ s-Ontak construct (SEQ ID NOs: 58-59; “His₆” disclosed as SEQ ID NO: 23). Specifically, a recombinant *C. diphtheriae* harboring a gene construct encoding s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was grown to OD ~12. The culture supernatant was harvested, concentrated by tangential flow filtration using a 30 kDa molecular weight cut-off membrane, and diafiltered for buffer exchange using tangential flow filtration as above. The protein mixture was partially purified by Ni-affinity chromatography and then purified to greater than 97% by gel permeation chromatography using S-100 resin. The resulting s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was >97% pure and stable at 4° C.

[0041] FIG. 18 illustrates purified VLM s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) at greater than 97% purity produced using the C-terminal His₆ VLM s-Ontak construct (SEQ ID NO: 23). Specifically, a recombinant *C. diphtheriae* harboring a gene construct encoding VLM s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was grown to OD ~12. The culture supernatant was harvested, concentrated by tangential flow filtration using a 30 kDa molecular weight cut-off membrane, and diafiltered for buffer exchange using tangential flow filtration as above. The protein mixture was partially purified by Ni-affinity chromatography and then purified to greater than 97% by gel permeation chromatography using S-100 resin. The resulting VLM s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was >97% pure.

[0042] FIG. 19 illustrates the S-100 gel filtration column used to purify s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) and VLM s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was calibrated for retention of proteins of known molecular weight: g-globulin (158 kDa), ovalbumin (43.5 kDa), and myoglobin (17 kDa). The retention time for s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) was 94 minutes, confirming that the s-Ontak-His₆ polypeptide is a >97% aggregate-free, full-length, monomeric diphtheria toxin fusion protein with an apparent molecular weight of 58 kDa and neither dimers nor higher order aggregates were detected by immunoblot probed with monoclonal anti-IL-2 antibody.

DEFINITIONS

[0043] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with

a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

[0044] The term “activity” refers to the ability of a gene to perform its function such as Indoleamine 2,3-dioxygenase (an oxidoreductase) catalyzing the degradation of the essential amino acid tryptophan (trp) to N-formyl-kynurenine.

[0045] By “agent” is meant any small molecule chemical compound, antibody, nucleic acid molecule, or polypeptide, or fragments thereof.

[0046] By “ameliorate” is meant decrease, suppress, attenuate, diminish, arrest, or stabilize the development or progression of a disease.

[0047] By “alteration” is meant a change (increase or decrease) in the expression levels or activity of a gene or polypeptide as detected by standard art known methods such as those described herein. As used herein, an alteration includes a 10% change in expression levels, preferably a 25% change, more preferably a 40% change, and most preferably a 50% or greater change in expression levels.

[0048] By “analog” is meant a molecule that is not identical, but has analogous functional or structural features. For example, a polypeptide analog retains the biological activity of a corresponding naturally-occurring polypeptide, while having certain biochemical modifications that enhance the analog’s function relative to a naturally occurring polypeptide. Such biochemical modifications could increase the analog’s protease resistance, membrane permeability, or half-life, without altering, for example, ligand binding. An analog may include an unnatural amino acid.

[0049] c- means “classic” when attached to a term such as c-denileukin diftotox means Ontak® or that commercially available protein.

[0050] By “disease” is meant any condition or disorder that damages or interferes with the normal function of a cell, tissue, or organ. Examples of diseases include cancer and tuberculosis.

[0051] By “effective amount” is meant the amount of a required to ameliorate the symptoms of a disease relative to an untreated patient. The effective amount of active compound(s) used to practice the present invention for therapeutic treatment of a disease varies depending upon the manner of administration, the age, body weight, and general health of the subject. Ultimately, the attending physician or veterinarian will decide the appropriate amount and dosage regimen. Such amount is referred to as an “effective” amount.

[0052] The term “express” refers to the ability of a gene to express the gene product including for example its corresponding mRNA or protein sequence(s).

[0053] By “fragment” is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, preferably, at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

[0054] is—means “immature secreted” when attached to a term such as is-denileukm diftotox means immature secreted denileukin diftotox that contains a signal peptide.

[0055] n—means “mature secreted” when attached to a term such as ms-denileukin diftotox means mature secreted denileukin diftotox that has been processed and no longer contains a signal peptide.

[0056] n—means “new” when attached to a term such as n-denileukin diftotox means new denileukin diftotox.

[0057] As used herein, the terms “prevent,” “preventing,” “prevention,” “prophylactic treatment” and the like refer to reducing the probability of developing a disorder or condition in a subject, who does not have, but is at risk of or susceptible to developing a disorder or condition.

[0058] The terms “polypeptide,” “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Polypeptides can be modified, e.g., by the addition of carbohydrate residues to form glycoproteins. The terms “polypeptide,” “peptide” and “protein” include glycoproteins, as well as non-glycoproteins.

[0059] The term “purity” refers to the amount of polypeptide of the invention present in a pharmaceutical composition free of other polypeptides. For example, a polypeptide of the invention present in a pharmaceutical composition having a purity of about 80% means that greater than about 80% of polypeptide is full-length and contaminated by less than about 20% of either product-related or unrelated polypeptides. Purity can be determined, for example, by SDS polyacrylamide gel electrophoresis and staining with Coomassie blue, methods which are described in this application or by other methods known to those skilled in the art.

[0060] The term “aggregate-free, full-length, monomeric polypeptide” refers to the amount of polypeptide of the invention present in a pharmaceutical composition in monomeric form. For example, a pharmaceutical composition of the invention comprising greater than about 80% aggregate-free, full-length, monomeric polypeptide means that greater than about 80% of the full-length polypeptide is in monomeric form. The amount of aggregate-free, full-length, monomeric polypeptide can be determined, for example, by gel permeation chromatography using known monomeric polypeptides as size standards or by non-reducing, SDS-free native polyacrylamide gel electrophoresis, methods which are described in this application or by other methods known to those skilled in the art.

[0061] By “reduces” is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

[0062] A “reference” refers to a standard or control conditions such as a sample (human cells) or a subject that is a free, or substantially free, of an agent such as one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof

[0063] A “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset of or the entirety of a specified sequence; for example, a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. For polypeptides, the length of the reference polypeptide sequence will generally be at least about 16 amino acids, preferably at least

about 20 amino acids, more preferably at least about 25 amino acids, and even more preferably about 35 amino acids, about 50 amino acids, or about 100 amino acids. For nucleic acids, the length of the reference nucleic acid sequence will generally be at least about 50 nucleotides, preferably at least about 60 nucleotides, more preferably at least about 75 nucleotides, and even more preferably about 100 nucleotides or about 300 nucleotides or any integer thereabout or there between.

[0064] s—means “secreted” when attached to a term such as s-denileukin diftotox means secreted denileukin diftotox. Secreted denileukin diftotox includes is- and m-forms.

[0065] As used herein, the term “subject” is intended to refer to any individual or patient to which the method described herein is performed. Generally the subject is human, although as will be appreciated by those in the art, the subject may be an animal. Thus other animals, including mammals such as rodents (including mice, rats, hamsters and guinea pigs), cats, dogs, rabbits, farm animals including cows, horses, goats, sheep, pigs, etc., and primates (including monkeys, chimpanzees, orangutans and gorillas) are included within the definition of subject.

[0066] By “substantially identical” is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). Preferably, such a sequence is at least 60%, more preferably 80% or 85%, and more preferably 90%, 95% or even 99% identical at the amino acid level or nucleic acid to the sequence used for comparison.

[0067] Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

[0068] Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

[0069] As used herein, the terms “treat,” “treating,” “treatment,” and the like refer to reducing or ameliorating a disorder and/or symptoms associated therewith. It will be appreciated that, although not precluded, treating a disorder or condition does not require that the disorder, condition or symptoms associated therewith be completely eliminated.

[0070] Unless specifically stated or obvious from context, as used herein, the term “or” is understood to be inclusive.

Unless specifically stated or obvious from context, as used herein, the terms “a,” “an,” and “the” are understood to be singular or plural.

[0071] Unless specifically stated or obvious from context, as used herein, the term “about” is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

[0072] The recitation of a listing of chemical groups in any definition of a variable herein includes definitions of that variable as any single group or combination of listed groups. The recitation of an embodiment for a variable or aspect herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

[0073] Any compositions or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

[0074] VLM—means “vascular leakage mutant” when attached to a term such as denileukin diftotox-VLM means denileukin diftotox vascular leakage mutant.

[0075] w—means “wild type” when attached to a term such as w-diphtheria toxin means wild type-diphtheria toxin.

DETAILED DESCRIPTION OF THE INVENTION

[0076] One embodiment of the present invention is the discovery of a process that produces aggregate-free, monomeric, diphtheria toxin fusion proteins having enhanced purity and quality.

[0077] This process includes transforming bacteria including preferably, strains of *Corynebacterium diphtheria* with DNA expression vectors of the present invention. DNA expression vectors of the present invention are designed to include specific genetic elements comprising a tox promoter (toxP) and an overlapping novel, mutated tox operator (toxO), preferably a signal sequence, and a DNA sequence encoding a protein. The protein is preferably a fusion protein including a diphtheria toxin, or functional part thereof, and a target receptor binding domain or a functional part thereof. The term “functional part thereof” means a part of a diphtheria toxin protein that acts as a toxin or the part of a target receptor binding domain that binds to its receptor. DNA expression vectors of the present invention are designed so proteins are expressed from a tox promoter (toxP) and a mutant tox operator (toxO).

Mutant toxO

[0078] toxO, is a 19-bp operator region that is composed of two 9 bp imperfect palindromic arms interrupted by a central cytosine (C) base. The wild type toxO (FIG. 1b) and a mutant toxO (FIG. 1a) operator discovered by inventors are shown in FIG. 1. SEQ ID NO: 1 illustrates one embodiment of the DNA sequence of a mutant toxO this invention. toxP is a promoter having a DNA sequence of SEQ ID NO: 2. SEQ ID NO: 2 illustrates the toxP DNA sequences include the toxO DNA sequences. SEQ ID NO: 3 is a DNA sequence including a toxP, a toxO, a signal peptide, and a DNA sequence encoding a protein. The asterisks in SEQ ID NO: 3 indicate the changes introduced to create the mutant toxO.

(Mutant toxO DNA sequence)

SEQ ID NO: 1

TTAGGATAGCTAAGTCCAT

(toxP including the mutant toxO DNA sequence
where the mutant toxO sequence is underlined)

SEQ ID NO: 2

TTGATTTTCAGAGCACCTTATAATTAGGATAGCTAAGTCCAT

[0079] The toxO DNA operator sequence is bound by a protein known as the diphtheria toxin repressor, DtxR. DtxR is a global iron-activated regulatory protein that is able to control gene expression. In iron-replete conditions, Fe^{2+} and Fe^{3+} ions bind to apo-DtxR causing a conformational change that allows the formation of homodimers of the DtxR repressor, which bind to the tox operator (toxO) DNA sequence and repress tox gene expression. In low iron environments, Fe^{2+} and Fe^{3+} ions disassociate from DtxR causing it to lose its DNA binding capability and disassociate from the operator; this event thereby allows expression of tox gene products. FIG. 1b illustrates the wild type toxO DNA sequence.

[0080] To overcome the inhibitory effect of Fe^{2+} and Fe^{3+} ions on tox expression, a DNA expression vector was created replacing the wild type (WT) toxO with a mutant toxO DNA sequence. This change blocks Fe ion-mediated regulation of tox gene expression. FIG. 1a, SEQ ID NO: 1, and SEQ ID NO: 3 illustrate the mutant /OA) DNA sequence of the present invention. Under this invention, bacteria such as *E. coli* and *C. diphtheria* harboring a recombinant plasmid encoding a diphtheria toxin fusion protein under the control of toxP and the mutant toxO may be grown in Fe-replete media, allowed to grow to high densities, and will not require a shift to Fe-free media to induce expression. The constitutive expression of tox gene products in iron replete medium represents a significant advance in the field. *C. diphtheria*, specifically the C7 beta (-), tox (-) strain is the preferred host bacteria for the production of all diphtheria-toxin related recombinant proteins using the DNA expression vectors of the present invention. The DNA expression vectors of the present invention may be used in other bacteria such as *E. coli*.

DNA Expression Vectors

[0081] The DNA expression vectors of the present invention includes a toxP, mutant toxO, a DNA sequence encoding a protein, and preferably a signal sequence. SEQ ID NO: 3 is one example of a DNA sequence containing these genetic elements that may be part of a DNA expression vector of the present invention. As mentioned, the asterisks observed in SEQ ID NO: 3 are placed above the base pair changes between the mutant and wild type toxO. SEQ ID NO: 3 is numbered such that the toxP extends from base 1 to 30, and toxO begins at base 24 and ends at base 42 (prior to the underlined DNA sequence). The underlined DNA sequence represents base 74 to base 148 and is a region of DNA encoding a 25 amino acid signal sequence (also observe in SEQ ID NO: 4, SEQ ID NO: 5, and FIG. 2). The DNA expression vectors of the present invention are preferably constructed so one or more proteins are expressed

from toxP, mutant toxO, and are translated with an N-terminal signal sequence. The N-terminal signal sequence targets the one or more proteins (expressed from the vector) for secretion, and the N-terminal signal peptide is later cleaved to make mature active proteins. SEQ ID NO: 3 includes DNA sequences encoding proteins such as a novel denileukin difttox called secreted-denileukin difttox, or s-denileukin difttox. The s-denileukin difttox has two forms called immature secreted-denileukin difttox (is-denileukin difttox) and mature secreted-denileukin difttox (ms-denileukin difttox). SEQ ID NO: 12 is of is-denileukin difttox of the present invention and SEQ ID NO: 13 is of ms-denileukin difttox of the present invention. The is-denileukin difttox contains a signal sequence that during processing is cleaved off to form ms-denileukin difttox. In addition, SEQ ID NO: 3 includes a DNA sequence beginning at base 149 to 1711 that encodes a protein, specifically a fusion protein containing the functional parts of a diphtheria toxin and the functional parts of IL 2. A new denileukin difttox fusion protein sequence is formed called ms-denileukin difttox that is a 520 amino acid polypeptide and is composed of the amino acid sequences for diphtheria toxin fragments A and a portion of fragment B (Gly₁-His₃₈₇) and the sequences for human interleukin-2. As a result of cleavage of the signal sequence, ms-denileukin difttox of the present invention lacks the first methionine present in classic-denileukin difttox (c-denileukin difttox) and is thereby one amino acid shorter than the amino acid sequence of the classic-denileukin difttox protein known as Ontak®. SEQ ID NO: 13 is the protein sequence of the new difttox protein sequence ms-denileukin difttox which may be compared to SEQ ID NO: 10 containing the protein sequence of the classic-denileukin difttox (c-denileukin difttox) known as Ontak®.

[0082] DNA expression vectors of the present invention include DNA sequences encoding one or more protein(s). A preferred protein of the present invention is a fusion protein comprising a diphtheria toxin (or a functional part thereof) and a target receptor binding protein (or a functional part thereof). An example of a diphtheria toxin that may be produced from a DNA expression is any functional part of a diphtheria toxin or any functional part of a diphtheria toxin vascular leakage mutant. Examples of proteins of target receptor binding domains produced from a DNA expression vector of the present invention include, IL-2, IL-3, IL-4, IL-6, IL-7, IL-15, EGF, FGF, substance P, CD4, α MSH, GRP, TT fragment C, GCSF, heregulin β 1, TNF α , TGF β , or a combination thereof. Other target receptor binding domains may be used depending upon the therapeutic application; however, SEQ. ID NO. 9 is a preferred DNA sequence encoding a functional part of IL2 receptor binding domain. For the purposes of the present invention, some of the DNA plasmids and the genetic elements thereof are illustrated in FIG. 1, FIG. 2, FIG. 6, and FIG. 7. Examples of fusion proteins encoded by DNA expression vectors of the present invention include SEQ ID NOs: 11, 12, 13, 14, 15, 19, and 21.

(DNA sequence encoding secreted-denileukin diftitox or s-denileukin diftitox Sequence includes toxP, mutant tox0, signal sequence, a functional part of diphtheria toxin and a functional part of IL2. Bold font and asterisks indicate the changes introduces to create the mutant tox0).

SEQ ID NO: 3

1 TTGATTTTCAGAGCACCCCTTATAATTAGGATAGCT**AAGTCCATT**ATTTTAT

51 GAGTCCTGGTAAGGGGATACGTTGTGAGCAGAA**AACTGTTTGCGTCAATC**

101 TTAATAGGGGCGCTACTGGGGATAGGGGCCCCACCTTCAGCCCATGCAGG

151 CGCTGATGATGTTGTTGATTCTTCTAAATCTTTTGATGGAA**AACTTTT**

201 CTTTCGTACCACGGGACTAAACCTGGTTATGTAGATTCCATTCAAAAAGGT

251 ATACAAAAGCCAAAATCTGGTACACAAGGAAATTATGACGATGATTGGAA

301 AGGGTTTTATAGTACCGACAATAAATACGACGCTGCGGGATACTCTGTAG

351 ATAATGAAAACCCGCTCTCTGGAAAAGCTGGAGGCGTGGTCAAAGTGACG

401 TATCCAGGACTGACGAAGGTTCTCGCACTAAAAGTGGATAATGCCGAAAC

451 TATTAAGAAAGAGTTAGGTTTAAGTCTCACTGAACCGTTGATGGAGCAAG

501 TCGGAACGGAAGAGTTTATCAAAAGGTTTCGGTGATGGTGCTTCGCGTGTA

551 GTGCTCAGCCTTCCCTTCGCTGAGGGGAGTTCTAGCGTTGAATATATTAA

601 TAACTGGGAACAGGCGAAAGCGTTAAGCGTAGAACTTGAGATTAATTTTG

651 AAACCCGTGGAAAACGTGGCCAAGATGCGATGTATGAGTATATGGCTCAA

701 GCCTGTGCAGGAAATCGTGTACGGCGATCAGTAGGTAGCTCATTGTCATG

751 CATCAACCTGGATTGGGATGTTATCCGTGATAAACTAA**AACTAAGATCG**

801 AATCTCTGAAAGAACACGGTCCGATCAAAACAAAATGAGCGAAAGCCCG

851 AACAA**AACTGTATCTGAAGAAAAGCTAAACAGTACCTGGAAGAATTCCA**

901 CCAGACTGCACTGGAACACCCGGAACGTCTGAAC**T**TAAAGACCGTTACTG

951 GTACCAACCCGGTATTTCGCTGGTGCTAACTACGCTGCTTGGGCAGTAAAC

1001 GTTGCTCAGGTTATCGATAGCGAAACTGCTGATAACCTGGAAAAAACTAC

1051 CGCGGCTCTGTCTATCCTGCCGGGTATCGGTAGCGTAATGGGCATCGCAG

1101 ACGGCGCCGTTCAACCAACACTGAAGAAATCGTTGCACAGTCTATCGCT

1151 CTGAGCTCTCTGATGGTTGCTCAGGCCATCCCGCTGGTAGGTGAAC**TGGT**

1201 TGATATCGGTTTCGCTGCATACA**ACTTCGTTGAAAGCATCATCAACCTGT**

1251 TCCAGGTTGTTCA**CAACTCTTACAACCGCCCGCTTACTCTCCGGGTCAC**

1301 AAGACGCATGCACCTACTTCTAGCTCTACCAAGAAAACCCAGCTGCAGCT

1351 CGAGCACCTGCTGCTGGATTTGCAGATGATCCTGAACGGTATCAACAATT

1401 ACAAGAACCCGAAACTGACGCGTATGCTGACCTTCAAGTTCTACATGCCG

1451 AAGAAGGCCACCGAACTGAAACACCTGCAGTGTCTAGAAGAAGAACTGAA

1501 ACCGCTGGAGGAAGTTCTGAACCTGGCTCAGTCTAA**AACTTCCACCTGC**

1551 GGCCGCGTGACCTGATCTCTAACATCAACGTAATCGTTCTGGAAC**TGAAG**

1601 GGCTCTGAAACCACCTTCATGTGTGAATACGCTGATGAGACCGCAACCAT

1651 CGTAGAATTCCTGAACCGTTGGATCACCTTCTGTCAGTCTATCATCTCTA

1701 CCCTGACCTGA < 1711

-continued

(Signal DNA Sequence)SEQ ID NO: 4

74 GTGAGCAGAAAAC TGTTT GCGTCAATCTTAATAGGGGCGCTACTGGGGAT

124 AGGGGCCCCACCTTCAGCCCATGCA < 148

(Signal Protein Sequence)SEQ ID NO: 5

-25 MSRKLFASILIGALLGIGAPPSAHA < -1

(classic-denileukin diftitox DNA sequence)SEQ ID NO: 6

1 ATG

4 GCGCTGATGATGTTGTTGATTCTTCTAAATCTTTGTGATGGAAAAC TT

54 TTCTTCGTACCACGGGACTAAACCTGGTTATGTAGATTCCATTCAAAAAG

104 GTATACAAAAGCCAAAATCTGGTACACAAGGAAATTATGACGATGATTGG

154 AAAGGGTTTTATAGTACCGACAATAAATACGACGCTGCGGGATACTCTGT

204 AGATAATGAAAACCCGCTCTCTGGAAAAGCTGGAGGCGTGGTCAAAGTGA

254 CGTATCCAGGACTGACGAAGGTTCTCGCACTAAAAGTGGATAATGCCGAA

304 ACTATTAAGAAAGAGTTAGGTTTAAGTCTCACTGAACCGTTGATGGAGCA

354 AGTCGGAACGGAAGAGTTTATCAAAAGGTTCCGTGATGGTGCTTCGCGTG

404 TAGTGCTCAGCCTTCCCTTCGCTGAGGGGAGTTCTAGCGTTGAATATATT

454 AATAACTGGGAACAGGCGAAAGCGTTAAGCGTAGAACTTGAGATTAATTT

504 TGAAACCCGTGGAAAACGTGGCCAAGATGCGATGTATGAGTATATGGCTC

554 AAGCCTGTGCAGGAAATCGTGTCAGGCGATCAGTAGGTAGCTCATTGTCA

604 TGCATCAACCTGGATTGGGATGTTATCCGTGATAAAACTAAACTAAGAT

654 CGAATCTCTGAAAGAACACGGTCCGATCAAAAACAAAATGAGCGAAAGCC

704 CGAACAAAAC TGTATCTGAAGAAAAGCTAAACAGTACCTGGAAGAATTC

754 CACCAGACTGCACTGGAACACCCGGAACGTCTGAAC TTAAGACCGTTAC

804 TGGTACCAACCCGGTATTCGCTGGTGCTAACTACGCTGCTTGGGCAGTAA

854 ACGTTGCTCAGGTTATCGATAGCGAAACTGCTGATAACCTGGAAAAAACT

904 ACCGCGGCTCTGTCTATCCTGCCGGGTATCGGTAGCGTAATGGGCATCGC

954 AGACGGCGCCGTTCAACACAACACTGAAGAAATCGTTGCACAGTCTATCG

1004 CTCTGAGCTCTCTGATGGTTGCTCAGGCCATCCCGCTGGTAGGTGAACTG

1054 GTTGATATCGGTTTCGCTGCATACAAC TTCGTTGAAAGCATCATCAACCT

1104 GTTCCAGGTTGTTCAAACTCTTACAACCGCCCGGCTTACTCTCCGGGTC

1154 ACAAGACGCATGCACCTACTTCTAGCTCTACCAAGAAAACCCAGCTGCAG

1204 CTCGAGCACCTGCTGCTGGATTTGCAGATGATCCTGAACGGTATCAACAA

1254 TTACAAGAACCCGAAACTGACGCGTATGCTGACCTTCAAGTTCTACATGC

1304 CGAAGAAGGCCACCGAACTGAAACACCTGCAGTGTCTAGAAGAAGAACTG

1354 AAACCGCTGGAGGAAGTTCTGAACCTGGCTCAGTCTAAAAACTTCCACCT

1404 GCGGCCGCGTGACCTGATCTCTAACATCAACGTAATCGTTCTGGAAC TGA

1454 AGGGCTCTGAAACCACCTTCATGTGTGAATACGCTGATGAGACCGCAACC

1504 ATCGTAGAATTCCTGAACCGTTGGATCACCTTCTGTCAGTCTATCATCTC

1554 TACCCTGACCTGA < 1566

Formation of Diphtheria Toxin Fusion Proteins Having Minimal, or no, Vascular Leakage (Denileukin Difitox-VLMs)

[0083] Like all of the bacterial and plant toxins, denileukin difitox carries amino acid motifs that may induce vascular leak syndrome (VLS). Approximately 30% of patients treated with Ontak® develop VLS ranging from rapid weight gain with peripheral edema to hypoalbuminemia to pulmonary edema. Mutations were made to the DNA sequence of Ontak® as described in U.S. Pat. No. 8,865,866. It was discovered that DNA mutations made to the DNA sequence such that the valine (GTT) at the 7th residue of SEQ ID NO: 10 is replaced with an alanine as shown in SEQ ID NO: 16, resulted in the fusion toxin having little, or no, vascular leak syndrome side effects. These mutants are referred to as “vascular leak mutants” (VLM). The vascular leak mutants, or denileukin difitox-VLMs are shown to have the same potency as c-denileukin difitox in FIG. 3, not to cause vascular leak in FIG. 4, and to have significantly less acute toxicity in vivo than c-denileukin difitox in FIG. 5. s-denileukin difitox-VLM, has an alanine replacing the valine at the 6th residue shown in in SEQ ID NOs: 14 and 15. s-denileukin difitox-VLM protein should have a similar decrease in toxicity as that found with the c-denileukin difitox-VLM protein.

[0084] Also, the sequences V₂₉D₃₀S₃₁ and I₂₉₀D₂₉₁S₂₉₂ shown in SEQ ID NO: 10 (amino acid sequence of c-denileukin difitox), when mutated also will reduce VLS. A claim in this discovery is that introduction of substitutions in V₂₉D₃₀S₃₁ and/or I₂₉₀D₂₉₁S₂₉₂ such as V29A or I290A may be introduced into the corresponding positions of diphtheria toxin fusion proteins and that these substitutions will also have value in further reducing vascular leakage syndrome.

(denileukin difitox-VLM underlined codon encodes for alanine, here shown as GCT, described in U.S. Pat. No. 8,865,866.)

	SEQ ID NO: 7
1	ATG
4	GGCGCTGATGATGTT <u>GCT</u> GATTCTTCTAAATCTTTGTGATGGA AAACTT
54	TTCTTCGTACCACGGGACTAAACCTGGTTATGTAGATTCCATTC AAAAAG
104	GTATACAAAAGCCAAAATCTGGTACACAAGGAAATTATGACGAT GATTGG
154	AAAGGGTTTTATAGTACCGACAATAAATACGACGCTGCGGGATA CTCTGT
204	AGATAATGAAAACCCGCTCTCTGGAAAAGCTGGAGGCGTGGTCA AAGTGA
254	CGTATCCAGGACTGACGAAGGTTCTCGCACTAAAAGTGGATAAT GCCGAA
304	ACTATTAAGAAAGAGTTAGGTTTAAGTCTCACTGAACCGTTGAT GGAGCA
354	AGTCGGAACGGAAGAGTTTATCAAAGGTTGCGGTGATGGTGCTT CGCGTG

-continued

404	TAGTGCTCAGCCTTCCCTTCGCTGAGGGGAGTTCTAGCGTTGAA TATATT
454	AATAACTGGGAACAGGCGAAAGCGTTAAGCGTAGAACTTGAGAT TAATTT
504	TGAAACCCGTGGAAAACGTGGCCAAGATGCGATGTATGAGTATA TGGCTC
554	AAGCCTGTGCAGGAAATCGTGTGAGGCGATCAGTAGGTAGCTCA TTGTCA
604	TGCATCAACCTGGATTGGGATGTTATCCGTGATAAACTAAAC TAAGAT
654	CGAATCTCTGAAAGAACACGGTCCGATCAAAAACAAATGAGCG AAAGCC
704	CGAACAAACTGTATCTGAAGAAAAGCTAAACAGTACCTGGAA GAATTC
754	CACCAGACTGCACTGGAACACCCGGAAGTGTCTGAACCTAAGAC CGTTAC
804	TGGTACCAACCCGGTATTCGCTGGTGCTAACTACGCTGCTTGGG CAGTAA
854	ACGTTGCTCAGGTTATCGATAGCGAACTGCTGATAACCTGGAA AAACT
904	ACCGCGGCTCTGTCTATCCTGCCGGGTATCGGTAGCGTAATGGG CATCGC
954	AGACGGCGCCGTTTACCACAACACTGAAGAAATCGTTGCACAGT CTATCG
1004	CTCTGAGCTCTCTGATGGTTGCTCAGGCCATCCCGCTGGTAGGT GAACTG
1054	GTTGATATCGGTTTTCGCTGCATACAACCTTCGTTGAAAGCATCAT CAACCT
1104	GTTCCAGGTTGTTTCACTCTTACAACCGCCCGGCTTACTCTC CGGGTC
1154	ACAAGACGCATGCACCTACTTCTAGCTCTACCAAGAAAACCCAG CTGCAG
1204	CTCGAGCACCTGCTGCTGGATTTCAGATGATCCTGAACGGTAT CAACAA
1254	TTACAAGAACCCGAAACTGACGCGTATGCTGACCTTCAAGTTCT ACATGC
1304	CGAAGAAGGCCACCGAACTGAAACACCTGCTGCAGTGTCTAGAA GAAGAA
1354	CTGAAACCGCTGGAGGAAGTTCTGAACCTGGCTCAGTCTAAAAA CTTCCA
1404	CCTGCGGCCGCGTGACCTGATCTCTAACATCAACGTAATCGTTC TGGAAC
1454	TGAAGGGCTCTGAAACCACCTTCATGTGTGAATACGCTGATGAG ACCGCA
1504	ACCATCGTAGAATTCCTGAACCGTTGGATCACCTTCTGTCTAGTC TATCAT
1554	CTCTACCCTGACC < 1566

[0085] Alignment of DNA sequences comparing SEQ ID NO: 7 [denileukin difitox-VLM described in US patent No. 8,865,866] with SEQ ID NO: 8 [is-denileukin difitox-VLM of the present invention] demonstrates SEQ ID NO: 8 is missing a codon (three bases) in line 1381-1437.

Similarity: 1563/1638 (95.42%)			
NO: 7	1	-----	0
#####			
NO: 8	1	GTGAGCAGAAAAGTGTTCGCTCAATCTTAATAGGGGCGCTACTGGGGATAGGGGCCCA	60
NO: 7	1	-----ATG--GGCGCTGATGATGTTGCTGATTCTTCTAAATCTTTTGTGATGGAA	48
##### ##			
NO: 8	61	CCTTCAGCCCATGCAGGCGCTGATGATGTTGCTGATTCTTCTAAATCTTTTGTGATGGAA	120
NO: 7	49	AACTTTCTTCGTACCACGGGACTAAACCTGGTTATGTAGATTCCATTCAAAAAGGTATA	108
NO: 8	121	AACTTTCTTCGTACCACGGGACTAAACCTGGTTATGTAGATTCCATTCAAAAAGGTATA	180
NO: 7	109	CAAAAGCCAAAATCTGGTACACAAGGAAATTATGACGATGATTGGAAAGGGTTTATAGT	168
NO: 8	181	CAAAAGCCAAAATCTGGTACACAAGGAAATTATGACGATGATTGGAAAGGGTTTATAGT	240
NO: 7	169	ACCGACAATAAATACGACGCTGCGGGATACTCTGTAGATAATGAAAACCCGCTCTCTGGA	228
NO: 8	241	ACCGACAATAAATACGACGCTGCGGGATACTCTGTAGATAATGAAAACCCGCTCTCTGGA	300
NO: 7	229	AAAGCTGGAGGCGTGGTCAAAGTGACGTATCCAGGACTGACGAAGGTTCTCGCACTAAAA	288
NO: 8	301	AAAGCTGGAGGCGTGGTCAAAGTGACGTATCCAGGACTGACGAAGGTTCTCGCACTAAAA	360
NO: 7	289	GTGGATAATGCCGAAACTATTAAGAAAGAGTTAGGTTTAAGTCTCACTGAACCGTTGATG	348
NO: 8	361	GTGGATAATGCCGAAACTATTAAGAAAGAGTTAGGTTTAAGTCTCACTGAACCGTTGATG	420
NO: 7	349	GAGCAAGTCGGAACGGAAGAGTTTATCAAAAGGTTCCGGTGATGGTGCTTCGCGTGTAGTG	408
NO: 8	421	GAGCAAGTCGGAACGGAAGAGTTTATCAAAAGGTTCCGGTGATGGTGCTTCGCGTGTAGTG	480
NO: 7	409	CTCAGCCTTCCCTTCGCTGAGGGGAGTTCTAGCGTTGAATATATTAATAACTGGGAACAG	468
NO: 8	481	CTCAGCCTTCCCTTCGCTGAGGGGAGTTCTAGCGTTGAATATATTAATAACTGGGAACAG	540
NO: 7	469	GCGAAAGCGTTAAGCGTAGAACTTGAGATTAATTTTGAAACCCGTGGAAAACGTGGCCAA	528
NO: 8	541	GCGAAAGCGTTAAGCGTAGAACTTGAGATTAATTTTGAAACCCGTGGAAAACGTGGCCAA	600
NO: 7	529	GATGCGATGTATGAGTATATGGCTCAAGCCTGTGCAGGAAATCGTGTCAGGCGATCAGTA	588
NO: 8	601	GATGCGATGTATGAGTATATGGCTCAAGCCTGTGCAGGAAATCGTGTCAGGCGATCAGTA	660
NO: 7	589	GGTAGCTCATTGTTCATGCATCAACCTGGATTGGGATGTTATCCGTGATAAACTAAACT	648
NO: 8	661	GGTAGCTCATTGTTCATGCATCAACCTGGATTGGGATGTTATCCGTGATAAACTAAACT	720
NO: 7	649	AAGATCGAATCTCTGAAAGAACACGGTCCGATCAAAAACAAAATGAGCGAAAGCCCGAAC	708
NO: 8	721	AAGATCGAATCTCTGAAAGAACACGGTCCGATCAAAAACAAAATGAGCGAAAGCCCGAAC	780
NO: 7	709	AAACTGTATCTGAAGAAAAAGCTAAACAGTACCTGGAAGAATTCCACCAGACTGCACTG	768
NO: 8	781	AAACTGTATCTGAAGAAAAAGCTAAACAGTACCTGGAAGAATTCCACCAGACTGCACTG	840
NO: 7	769	GAACACCCGGAAGTGTCTGAACTTAAGACCGTTACTGGTACCAACCCGGTATTCGCTGGT	828
NO: 8	841	GAACACCCGGAAGTGTCTGAACTTAAGACCGTTACTGGTACCAACCCGGTATTCGCTGGT	900
NO: 7	829	GCTAACTACGCTGCTTGGGCAGTAAACGTTGCTCAGGTTATCGATAGCGAAACTGCTGAT	888
NO: 8	901	GCTAACTACGCTGCTTGGGCAGTAAACGTTGCTCAGGTTATCGATAGCGAAACTGCTGAT	960
NO: 7	889	AACCTGGAAAAAACTACCGCGGCTCTGTCTATCCTGCCGGGTATCGGTAGCGTAATGGGC	948
NO: 8	961	AACCTGGAAAAAACTACCGCGGCTCTGTCTATCCTGCCGGGTATCGGTAGCGTAATGGGC	1020
NO: 7	949	ATCGCAGACGGCGCCGTTTACCACAACACTGAAGAAATCGTTGCACAGTCTATCGCTCTG	1008
NO: 8	1021	ATCGCAGACGGCGCCGTTTACCACAACACTGAAGAAATCGTTGCACAGTCTATCGCTCTG	1080
NO: 7	1009	AGCTCTCTGATGGTTGCTCAGGCCATCCCGCTGGTAGGTGAACTGGTTGATATCGGTTTC	1068
NO: 8	1081	AGCTCTCTGATGGTTGCTCAGGCCATCCCGCTGGTAGGTGAACTGGTTGATATCGGTTTC	1140

- continued

NO: 7	1069	GCTGCATACAACCTTCGTTGAAAGCATCATCAACCTGTTCCAGGTTGTTTACAACTCTTAC	1128
NO: 8	1141	GCTGCATACAACCTTCGTTGAAAGCATCATCAACCTGTTCCAGGTTGTTTACAACTCTTAC	1200
NO: 7	1129	AACCGCCCGGCTTACTCTCCGGGTCACAAGACGCATGCACCTACTTCTAGCTCTACCAAG	1188
NO: 8	1201	AACCGCCCGGCTTACTCTCCGGGTCACAAGACGCATGCACCTACTTCTAGCTCTACCAAG	1260
NO: 7	1189	AAAACCCAGCTGCAGCTCGAGCACCTGCTGCTGGATTTGCAGATGATCCTGAACGGTATC	1248
NO: 8	1261	AAAACCCAGCTGCAGCTCGAGCACCTGCTGCTGGATTTGCAGATGATCCTGAACGGTATC	1320
NO: 7	1249	AACAATTACAAGAACCCGAAACTGACGCGTATGCTGACCTTCAAGTTCTACATGCCGAAG	1308
NO: 8	1321	AACAATTACAAGAACCCGAAACTGACGCGTATGCTGACCTTCAAGTTCTACATGCCGAAG	1380
NO: 7	1309	AAGGCCACCGAACTGAAACACCTGCTGCAGTGTCTAGAAGAAGAACTGAAACCGCTGGAG	1368
NO: 8	1381	AAGGCCACCGAACTGAAACACCTGC---AGTGTCTAGAAGAAGAACTGAAACCGCTGGAG	1437
NO: 7	1369	GAAGTTCTGAACCTGGCTCAGTCTAAAAACTTCCACCTGCGGCCGCTGACCTGATCTCT	1428
NO: 8	1438	GAAGTTCTGAACCTGGCTCAGTCTAAAAACTTCCACCTGCGGCCGCTGACCTGATCTCT	1497
NO: 7	1429	AACATCAACGTAATCGTTCTGGAAGTGAAGGGCTCTGAAACCACCTTCATGTGTGAATAC	1488
NO: 8	1498	AACATCAACGTAATCGTTCTGGAAGTGAAGGGCTCTGAAACCACCTTCATGTGTGAATAC	1557
NO: 7	1489	GCTGATGAGACCGCAACCATCGTAGAATTCCTGAACCGTTGGATCACCTTCTGTCAGTCT	1548
NO: 8	1558	GCTGATGAGACCGCAACCATCGTAGAATTCCTGAACCGTTGGATCACCTTCTGTCAGTCT	1617
NO: 7	1549	ATCATCTCTACCCTGACC---	1566
NO: 8	1618	ATCATCTCTACCCTGACCTGA	1638

(DNA sequence IL-2 portion of denileukin diftitox)		SEQ ID NO: 9
1	GCACCTACTTCTAGCTCTACCAAGAAAACCCAGCTGCAGCTCGAGCACCT	
51	GCTGCTGGATTTGCAGATGATCCTGAACGGTATCAACAATTACAAGAACC	
101	CGAAACTGACGCGTATGCTGACCTTCAAGTTCTACATGCCGAAGAAGGCC	
151	ACCGAACTGAAACACCTGCAGTGTCTAGAAGAAGAACTGAAACCGCTGGA	
201	GGAAGTTCTGAACCTGGCTCAGTCTAAAAACTTCCACCTGCGGCCGCGTG	
251	ACCTGATCTCTAACATCAACGTAATCGTTCTGGAAGTGAAGGGCTCTGAA	
301	ACCACCTTCATGTGTGAATACGCTGATGAGACCGCAACCATCGTAGAATT	
351	CCTGAACCGTTGGATCACCTTCTGTCTAGTCTATCATCTCTACCCTGACCT	
401	GA < 402	

Proteins Produced Using DNA Expression Vectors of the Present Invention

[0086] The first amino acid of a mature active diphtheria toxin related fusion protein of the present invention is a glycine as shown in bold (amino acid 1) in SEQ ID NOs: 13 and 15. The signal sequence within SEQ ID NO: 4 is labeled with negative numbers, counting back from the first glycine of the mature fusion protein and has the following amino acid sequence MSRKLFASILIGALLGIGAPPSAHA (SEQ ID NO: 22). The signal sequence is shown in SEQ ID NOs: 11 and 12 and is underlined. The mature secreted diphtheria toxin fusion protein includes a diphtheria toxin portion, such as Gly₁-His₃₈₇, and a target receptor binding domain, such

as an IL-2 protein from Ala388-Thr520 in SEQ ID NO: 3. Other target receptor binding domains used in the present invention that may be fused to a diphtheria toxin protein (or functional part thereof) include IL-3, IL-4, TL-6, IL-7, IL-15, EGF, FGF, substance P, CD4, αMSH, GRP, TT fragment C, GCSF, heregulin β1, TNFα, TGFβ, among others, or a combination thereof. SEQ ID NO: 10 describes c-denileukin diftitox that is not secreted and is requires purification from inclusion bodies in *E. coli*. SEQ ID NO: 12 describes immature secreted is-denileukm diftitox with a signal sequence. SEQ ID NO: 13 describes MS-denileukin diftitox wherein the signal sequence has been cleaved off during the process of secretion to the extracellular space.

(Protein Sequence of c-denileukin diftitox known as Ontak[®])

SEQ ID NO: 10

1 MGADDVVDSSKSFVMENFSSYHGTKP

27 GYVDSIQKGIQKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

77 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

127 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRKGQ

177 DAMYEYMAQACAGNRVRRSVGSSSLSCINLDWDVIRDKTKTKIESLKEHGP

227 IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

277 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

327 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

377 NRPAYSPGHKTHAPTSSSTKKTQLQLEHLLLDLQMLNGINNYKNPKLTR

427 MLTFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFHLRPRDLISN

477 INVIVLELKCSETTFMCEYADETATIVEFLNRWITFCQSIISTLT < 521

(w-diphtheria toxin)

SEQ ID NO: 11

1 MSRKLFASILIGALLGIGAPPSAHAGADDVVDSSKSFVMENFSSYHGTKP

51 GYVDSIQKGIQKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

101 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

151 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRKGQ

201 DAMYEYMAQACAGNRVRRSVGSSSLSCINLDWDVIRDKTKTKIESLKEHGP

251 IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

301 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

351 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

401 NRPAYSPGHKTQPFLHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENT

451 PLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRCRAIDGDVTFCRPKS

501 PVYVONGVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHQKVN

551 KLSLFFEIKS < 560

(is-denileukin diftitox)

SEQ ID NO: 12

-25 MSRKLFASILIGALLGIGAPPSAHAGADDVVDSSKSFVMENFSSYHGTKP

26 GYVDSIQKGIQKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

76 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

126 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRKGQ

176 DAMYEYMAQACAGNRVRRSVGSSSLSCINLDWDVIRDKTKTKIESLKEHGP

226 IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

276 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

326 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

376 NRPAYSPGHKTHAPTSSSTKKTQLQLEHLLLDLQMLNGINNYKNPKLTR

426 MLTFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFHLRPRDLISN

476 INVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT < 520

-continued

(ms-denileukin diftitox)

SEQ ID NO: 13

1 GADDVVDSSKSFVMENFSSYHGTPK

26 GYVDSIQKGIQKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

76 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

126 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRKGQ

176 DAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGP

226 IKNKMSESPNKTVSEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

276 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

326 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

376 NRPAYSPGHKTHAPTSSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTR

426 MLTFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFHLRPRDLISN

476 INVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT < 520

(Protein sequence of is-denileukin diftitox-VLM)

SEQ ID NO: 14

-25 MSRKLFASILIGALLGIGAPPSAHAGADDVADSSKSFVMENFSSYHGTPK

26 GYVDSIQKGIQKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

76 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

126 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRKGQ

176 DAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGP

226 IKNKMSESPNKTVSEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

276 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

326 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

376 NRPAYSPGHKTHAPTSSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTR

426 MLTFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFHLRPRDLISN

476 INVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT < 520

(Protein sequence of ms-denileukin diftitox-VLM)

SEQ ID NO: 15

1 GADDVADSSKSFVMENFSSYHGTPK

26 GYVDSIQKGIQKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

76 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

126 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRKGQ

176 DAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGP

226 IKNKMSESPNKTVSEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

276 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

326 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

376 NRPAYSPGHKTHAPTSSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTR

426 MLTFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFHLRPRDLISN

476 INVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT < 520

-continued

(Protein sequence of denileukin diftitox-VLM described in U.S. Pat. No. 8,865,866)

SEQ ID NO: 16

1 MGADDVADSSKSFVMENFSSYHGTKP

27 GYVDSIQKGIQKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

77 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

127 RFGDGASRVVLSLPPAEGSSSVEYINNWEQAKALSVELEINFETRGRGQ

177 DAMYEYMAQACAGNRVRRSVGSSSLSCINLDWDVIRDKTKTKIESLKEHGP

227 IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

277 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

327 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVHNSY

377 NRPAYSPGHKTHAPTSSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTR

427 MLTFKFYMPKKATELKHLQLCLEEELKPLEEVLNLAQSKNFHLRPRDLIS

477 NINVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT < 522

[0087] Protein Alignment of SEQ ID NO: 16 is denileukin diftitox-VLM described in U.S. Pat. No. 8,865,866 that has an extra amino acid (L) at position 445 when compared with SEQ ID NO: 14 is-denileukin diftitox-VLM of the present invention.

Use of DNA Expression Vectors to Manufacture Proteins.

[0088] The method using Fe-independent, secreted expression of proteins related to diphtheria toxin described above has several commercial applications in addition to the

Similarity: 521/522 (99.81%)					
NO: 16	1	M-----GADDVADSSKSFVMENFSSYHGTKPGYVDSIQKGI	36		
		#####			
NO: 14	1	MSRKLFASILIGALLGIGAPPSAHAGADDVADSSKSFVMENFSSYHGTKPGYVDSIQKGI	60		
NO: 16	37	QKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTYPGLTKVLALK	96		
NO: 14	61	QKPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGWKVTPGLTKVLALK	120		
NO: 16	97	VDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPPAEGSSSVEYINNWEQ	156		
NO: 14	121	VDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPPAEGSSSVEYINNWEQ	180		
NO: 16	157	AKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSSLSCINLDWDVIRDKTKT	216		
NO: 14	181	AKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSSLSCINLDWDVIRDKTKT	240		
NO: 16	217	KIESLKEHGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG	276		
NO: 14	241	KIESLKEHGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG	300		
NO: 16	277	ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIABGAVHHNTEEIVAQSIAL	336		
NO: 14	301	ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIAL	360		
NO: 16	337	SSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQWHNSYNRPAYSPGHKTHAPTSSSTK	396		
NO: 14	361	SSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQWHNSYNRPAYSPGHKTHAPTSSSTK	420		
NO: 16	397	KTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKFYMPKKATELKHLQLCLEEELKPLE	456		
NO: 14	421	KTQLQLEHLLLDLQMILNGINNYKNPKLTRMLTFKFYMPKKATELKHL-QCLEEELKPLE	479		
NO: 16	457	EVLNLAQSKNFHLRPRDLISNINVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQS	516		
NO: 14	480	EVLNLAQSKNFHLRPRDLISNINVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQS	539		
NO: 16	517	IISTLT	522		
NO: 14	540	IISTLT	545		

use of the method to express s-denileukin diftitox. The method can be used to improve (enhance) expression (yield) of:

WT Diphtheria Toxin:

[0089] The wild type Diphtheria toxin (SEQ ID NO: 11) used to make diphtheria toxoid, a vaccine for diphtheria which is present in DTP, TDaP, and other combination vaccines may be made using the DNA expression vector of the present invention. The DNA segment encoding SEQ ID NO: 11 may be placed in the DNA expression vector of the present invention and located downstream of the ToxP/ mutant ToxO.

Cross-Reacting Material-197 (CRM197) and Cross-Reacting Material-107 (CRM107):

[0090] CRM197 and CR107 are mutant proteins of full-length diphtheria toxin which are highly immunogenic but are completely devoid of toxin activity. They are used as carriers for several polysaccharide conjugate vaccines. For example, Wyeth and Pfizer took advantage of this immunogenicity in the 1990s when it conjugated seven polysaccharides from Streptococcus pneumoniae to CRM197 to create the original Prevnar vaccine which was FDA approved in February 2000. A 13-polysaccharides Prevnar was FDA-approved in 2010. The meningococcal vaccine Menveo, from Novartis, is four Neissena meningitidis polysaccharides plus CRM197. This vaccine gained FDA approval in

2010. The cancer immunotherapy company Imugene (ASX: IMU) reported dramatic improvements in antibody titers from its B cell peptide cancer immunotherapy targeting HER2 when it used CRM 197 as a carrier protein. CRM 197 is also being evaluated as a potential drug delivery protein. The Swiss-based Turing Pharmaceuticals is working on CRM197 fusion constructs with therapeutic proteins of up to 1,000 amino acids in length. The DNA expression vectors of the present invention maybe used to produce CRM 197 and CRM 107. One or more of the DNA segment(s) encoding SEQ ID NOs: 18-21 may be placed in the DNA expression vector of the present invention and located downstream of the ToxP/mutant ToxO.

Diphtheria Toxin Based Fusion Proteins with Cleavable Peptide or Protein Tags Used to Enhance Purification.

[0091] Cleavable peptide tags (such as His₆ (SEQ ID NO: 23) or FLAG [DYKDDDDK] (SEQ ID NO: 24)) or protein tags (such as GST [glutathione S-transferase] or SUMO [Small Ubiquitin-like Modifier protein]) may be fused with specific protease cleavage sites to diphtheria toxin based fusion proteins. Affinity chromatography methods using antibodies or ligands which bind to the tag may be used for rapid purification of the tagged protein.

[0092] Following purification, the specific cleavage site enables separation of the tag from the desired diphtheria toxin related proteins. Such fusions may enhance purification of diphtheria toxin based fusion proteins of the present invention.

(Protein sequence of ms-CRM197)		SEQ ID NO: 17
1	GADDVVDSSKSFVMENFSSYHGTPGYVDSIQKGIQPKSGTQGNYYDDDW	
51	KEFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTYPGLTKVLALKVDNAE	
101	TIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYI	
151	NNWEQAKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSLS	
201	CINLDWDVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSEEKAKQYLEEF	
251	HQTALEHPELSELKTVTGTNPVFAGANYAAWAVNVAQVIDSETADNLEKT	
301	TAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVGEL	
351	VDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPFLHDGYAVSWNT	
401	VEDSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVNKSKTHI	
451	SVNGRKIRMRCRAIDGDVTFCRPKSPVYVGNVHANLHVAFHRSSEKIH	
501	SNEISSDSIGVLGYQKTVDHSTKVNKSLSLFFEIKS	< 535
(Protein sequence of is-CRM197)		SEQ ID NO: 18
1	MSRKLFASILIGALLGIGAPPSAHAGADDVVDSSKSFVMENFSSYHGTP	
51	GYVDSIQKGIQPKSGTQGNYYDDDWKEFYSTDNKYDAAGYSVDNENPLSG	
101	KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK	
151	RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRGQ	
201	DAMYEYMAQACAGNRVRRSVGSSLS	CINLDWDVIRDKTKTKIESLKEHGP
251	IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG	
301	ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT	
351	EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY	

-continued

401 NRPAYSPGHKTQPFLHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENT

451 PLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRCRAIDGDVTFCRPKS

501 PVYVGNVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTKVNS

551 KLSLFFEIKS < 560

(Protein sequence of ms-CRM107)

SEQ ID NO: 19

GADDVVDSSKSFVMENFSSYHGTKP

51 GYVDSIQKGIQPKSGTQQNYDDDWKGFYSTDNKYDAAGYSVDNENPLSG

101 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

151 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRGQ

201 DAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGP

251 IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

301 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

351 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

401 NRPAYSPGHKTQPPFHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENT

451 PLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRCRAIDGDVTFCRPKS

501 PVYVGNVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTKVNF

551 KLSLFFEIKS < 560

(Protein sequence of is-CRM107)

SEQ ID NO: 20

1 MSRKLFASILIGALLGIGAPPSAHAGADDVVDSSKSFVMENFSSYHGTKP

51 GYVDSIQKGIQPKSGTQQNYDDDWKGFYSTDNKYDAAGYSVDNENPLSG

101 KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIK

151 RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRGQ

201 DAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGP

251 IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

301 ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

351 EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

401 NRPAYSPGHKTQPPFHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENT

451 PLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRCRAIDGDVTFCRPKS

501 PVYVGNVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTKVNF

551 KLSLFFEIKS < 560

TABLE 1	
SEQUENCE NUMBER	DESCRIPTION
SEQ ID NO: 38	Protein sequence of N terminal His tag to VLM s-Ontak
SEQ ID NO: 39	Protein sequence of N terminal His tag to VLM s-Ontak after signal sequence is cleaved
SEQ ID NO: 40	Protein sequence of N terminal His tag to VLM s-Ontak after signal sequence is cleaved and TEV site is cleaved
SEQ ID NO: 41	DNA sequence of N terminal His tag to VLM s-Ontak
SEQ ID NO: 42	Protein sequence of C terminal His tag to VLM s-Ontak
SEQ ID NO: 43	Protein sequence of C terminal His tag to VLM s-Ontak after signal sequence is cleaved)
SEQ ID NO: 44	DNA sequence of C terminal His tag to VLM s-Ontak

TABLE 1-continued

SEQUENCE NUMBER	DESCRIPTION
SEQ ID NO: 45	Protein sequence of C terminal TEV His9 tag to VLM s-Ontak (“His9” disclosed as SEQ ID NO: 48)
SEQ ID NO: 46	Protein sequence of C terminal TEV His9 tag to VLM s-Ontak after signal sequence is cleaved (“His9” disclosed as SEQ ID NO: 48)
SEQ ID NO: 30	Protein sequence of C terminal TEV His9 tag to VLM s-Ontak after signal sequence and Tev protease site are cleaved (“His9” disclosed as SEQ ID NO: 48)
SEQ ID NO: 31	DNA sequence of C terminal His tag to VLM s-Ontak
SEQ ID NO: 32	Secreted <i>C. diphtheriae</i> protease 1 amino acid sequence
SEQ ID NO: 33	Secreted <i>C. diphtheriae</i> protease 1 DNA sequence
SEQ ID NO: 34	DNA sequence of allelic exchange substrate [AES] for knocking out secreted <i>C. diphtheriae</i> protease 1
SEQ ID NO: 35	Secreted <i>C. diphtheriae</i> protease 2 amino acid sequence
SEQ ID NO: 36	Secreted <i>C. diphtheriae</i> protease 2 DNA sequence)
	Protease 2 DNA sequence
SEQ ID NO: 37	DNA sequence of allelic exchange substrate [AES] for knocking out secreted <i>C. diphtheriae</i> protease 2

Purification of VLM s-Ontak Using His-Tagged Versions of the Polypeptide

[0093] In some preparations of VLM s-Ontak produced in *Corynebacterium diphtheriae* C7 slow proteolytic cleavage of the mature 520 amino acid polypeptide occurs. This is probably due to secreted proteases made by *Corynebacterium diphtheriae* C7. This proteolytic cleavage occurs at approximately ammo acid 390 of the mature 520 ammo acid VLM s-Ontak.

[0094] Histidine-tagged (His-tagged) versions of VLM s-Ontak have been constructed for the purpose of accelerating the purification of the desired protein away from the secreted proteases present in the culture supernatant. Tobacco Etch Virus (TEV) nuclear-inclusion-a endopeptidase (EC 3.4.22.44) recognition sites have also been engineered into these His-tagged versions of VLM s-Ontak. The purpose of the TEV cleavage sites is to enable the removal of the poly-His sequences in the final preparation of VLM s-Ontak. TEV is a highly specific endopeptidase which recognizes the amino acid sequence ENLYFQX where ‘V’ denotes the cleaved peptide bond, and X represents any small hydrophobic or polar amino acid such as glycine (G) (SEQ ID NO: 49).

N-terminal His-Tagged VLM s-Ontak with TEV Cleavage Site.

[0095] As shown in SEQ ID: 38 (Protein sequence of N terminal His tag to VLM s-Ontak) it is possible to add the amino sequence HHHHHHENLYFQ (SEQ ID NO: 50) to the immature protein sequence of VLM s-Ontak near its N-termmus. In this version, the sequence HHHHHHENLYFQ (SEQ ID NO: 50) appears immediately after the 26 amino acid signal sequence and immediately before the mature sequence of VLM s-Ontak (GADDVA (SEQ ID NO: 51)). The first glycine of VLM s-Ontak comprises the final recognition residue for the TEV protease which recognizes ENLYFQX (SEQ ID NO: 49) with X being any small amino acid. The mature, secreted protein sequence of this N-terminal His-tagged VLM s-Ontak is shown in SEQ ID: 39 (Protein sequence of N terminal His tag to VLM s-Ontak after signal sequence is cleaved) which is a good candidate for Nickel-column affinity purification with its His₆ tag (SEQ ID NO: 23). The affinity purified VLM s-Ontak may then be exposed to small amounts of pure TEV protease, leading to enzymatic proteolysis that removes the 13 N-terminal residues MHHHHHHHENLYFQ (SEQ ID NO: 52) and

releases mature, untagged VLM s-Ontak as is shown in SEQ ID NO: 40 (Protein sequence of N terminal His tag to VLM s-Ontak after signal sequence is cleaved and TEV site is cleaved).

[0096] Because the secreted protease(s) of *Corynebacterium diphtheriae* C7 cleave at approximately amino acid 390, N-terminal His-tagging can lead to two species: full length desired VLM s-Ontak (520 amino acids) and a 390-amino acid N-termmal breakdown fragment. These two polypeptides, being relatively close in size (as well as molecular composition) are difficult to separate by size exclusion chromatography. Hence we have also developed C-terminal His-tagged version of VLM s-Ontak.

C-Terminal His-Tagged VLM s-Ontak Without TEV Cleavage Wite.

[0097] As shown in SEQ ID NO: 42 (Protein sequence of C terminal His tag to VLM s-Ontak) it is possible to add the amino sequence HHHHHH (SEQ ID NO: 23) to the immature protein sequence of VLM s-Ontak at its C-terminus. In this version, the sequence HHHHHH (SEQ ID NO: 23) appears immediately after the C-termmal threonine of VLM s-Ontak (. . . IISTLT (SEQ ID NO: 53)). The mature, secreted protein sequence of this C-terminal His-tagged VLM s-Ontak is shown in SEQ ID: 43 (Protein sequence of C terminal His tag to VLM s-Ontak after signal sequence is cleaved) which is a good candidate for Nickel-column affinity purification with its His₆ tag (SEQ ID NO: 23).

C-Terminal His-Tagged VLM s-Ontak with TEV Cleavage Site.

[0098] In order to avoid having the His₆ sequence (SEQ ID NO: 23) in the final polypeptide sequence of the above version of VLM s-Ontak made by C-terminal His-tagging (SEQ ID: 43), it is possible to insert a TEV recognition sequence at the C-terminus to enable removal of the His-tag sequence. In this version, the sequence ENLYFQGHHHHHHHHH (SEQ ID NO: 54) appears immediately after the C-terminal threonine of VLM s-Ontak (. . . IISTLT (SEQ ID NO: 53)). Since nickel affinity binding is enhanced by poly-His sequences even longer than six amino acids, it is possible to include nine His residues. The amino acid sequence of this C-terminal His-tagged VLM s-Ontak with TEV cleavage site is shown in SEQ ID: 45 (Protein sequence of C terminal TEV His9 tag (SEQ ID NO: 48) to VLM s-Ontak). The mature, secreted protein sequence of this C-terminal His-tagged VLM s-Ontak with TEV

cleavage site is shown in SEQ ID: 46 (Protein sequence of C terminal TEV His9 tag (SEQ ID NO: 48) to VLM s-Ontak after signal sequence is cleaved) and is a good candidate for Nickel-column affinity purification with its His⁹ tag (SEQ ID NO: 48). The affinity purified VLM s-Ontak may then be exposed to small amounts of pure TEV protease, leading to enzymatic proteolysis that removes the 10 C-terminal residues GHHHHHHHHH (SEQ ID NO: 55), and releases mature, untagged VLM s-Ontak as is shown in SEQ ID: 30. Of note, this version of purified VLM s-Ontak (SEQ ID: 30) is 526 amino acids in length rather than 520 amino acids (SEQ ID NO: 15) because it contains six additional amino acids of the TEV protease recognition sequence (ENLYFQ (SEQ ID NO: 56) fused to the usual C-terminus threonine of VLM s-Ontak (. . . IISTLT (SEQ ID NO: 53)). The end result of this version of C-terminal His-tagged VLM s-Ontak with TEV cleavage site (SEQ ID: 30) is a C-terminal sequence . . . IISTLTENLYFQ (SEQ ID NO: 57).

Manufacturing Method for VLM s-Ontak Which Include His-Tags and TEV Protease Sites.

[0099] The above three His-tag versions of VLM s-Ontak (N-terminal His₆ tag (SEQ ID NO: 23) with TEV protease site, C-terminal His₆ tag (SEQ ID NO: 23) without TEV protease site, and C-terminal His₉ (SEQ ID NO: 48) tag with TEV protease site) are examples of methods to use His-tag/Nickel column affinity chromatography in the manufacturing method of VLM s-Ontak. Because of secreted proteases from *Corynebacterium diphtheriae* C7 that are present in the culture supernatant, it is important to purify VLM s-Ontak away from other proteins in the culture supernatant rapidly in order to avoid significant loss of the desired product. The inclusion of His-tags and TEV protease sites represents a significant improvement and may enable a rapid, streamlined manufacturing process for VLM s-Ontak.

Generation of *Corynebacterium diphtheriae* C7 Lacking Key Secreted Proteases for Improved Manufacturing of VLM s-Ontak.

[0100] The genome sequence of *Corynebacterium diphtheriae* C7 reveals two secreted proteases: Protease 1 is NCBI Reference Sequence WP_014318592.1 (SEQ ID: 32, 33) and Protease 2 is NCBI Reference Sequence WP_014318898.1 (SEQ ID: 35, 36). These proteases may be genetically deleted using the method of Ton-That and Schneewind (Ton-That H, Schneewind O. Assembly of pili on the surface of *Corynebacterium diphtheriae*. Mol Microbiol. 2003 Nov;50(4): 1429-38. PubMed PMID: 14622427) and also Allen and Schmitt (Allen C E, Schmitt M P. HtaA is an iron-regulated hemin binding protein involved in the utilization of heme iron in *Corynebacterium diphtheriae*. J Bacteriol. 2009 Apr;191(8):2638-48. PubMed PMID: 19201805). The allelic exchange substrates to knock out protease 1 and protease 2 are shown in SEQ ID: 34 and SEQ ID: 37, respectively. These sequences when inserted into pk18mobsacB, a conjugative, mating plasmid with sacB counterselection (Schafer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, Pühler A (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145:69-73. PMID: 8045426), lead to constructs which will knockout each protease. A recombinant *Corynebacterium diphtheriae* strain lacking both protease 1 and protease 2 will be a valuable production strain for future manufacturing methods to generate VLM s-Ontak.

Protein Manufacturing Process of Diphtheria Toxin-Based Fusion Proteins

[0101] Using the DNA plasmids and expression vectors of the present invention, a novel process was discovered eliminating the problems associated with the conventional method of manufacturing Ontak®. Ontak® is currently expressed using a DNA vector in an *E. coli* expression system, c-denileukin diftotox or Ontak® is 521 amino acids in length and has a molecular weight of 58 kD. The conventional Ontak® manufacturing process results in the formation of Ontak® aggregates of heterogeneous molecular weight, residual DNA, and excessive residual detergent in the final formulation resulting in the FDA placing classic-Ontak® on clinical hold in June 2011. As observed in FIG. 8a, Ontak® is expressed from a plasmid in *E. coli* and results in insoluble, cytosolic Ontak® (protein) accumulations known as inclusion body forms. Using the process of the present invention, FIG. 8b illustrates the expression of s-denileukin diftotox as an extracellular mature secreted protein in a cell free supernatant that can be easily purified and results in higher protein yields as illustrated in FIG. 9. FIG. 9 shows both a Coomassie Blue stain for total protein and an anti-IL2 immunoblot of s-denileukin diftotox generated using the process of the present invention probed with anti-IL-2.

[0102] The novel process of the present invention comprises: 1) transforming bacteria, preferably a *Corynebacterium diphtheriae* strain, with a DNA expression vector of the present invention, 2) forming a transformant; 3) incubating the transformant in a culture medium for a period of time to allow growth and expression of a protein (such as a diphtheria toxin-based fusion protein and CRM typically containing a signal peptide), 4) secretion of the protein into the culture medium (due to a signal peptide attached to the protein); and (8) purifying the diphtheria toxin-based fusion protein from the culture medium. The DNA expression vectors include a ToxP and mutant ToxO that regulate the expression of at least one protein, such as a diphtheria toxin fusion protein, CRM protein, or other protein that may be attached to a signal peptide of the present invention.

Therapeutic Applications of Diphtheria Toxin-Based Fusion Proteins of the Present Invention

[0103] Clinical efficacy of Ontak® has been demonstrated in cutaneous T cell lymphoma, peripheral T cell lymphoma, steroid-refractory graft versus host disease, methotrexate-refractory psoriasis, and methotrexate-refractory rheumatoid arthritis. Clinical efficacy has also been demonstrated in malignant melanoma and ovarian carcinoma as shown in FIG. 14. The diphtheria toxin-based fusion proteins of the present invention (including s-denileukin diftotox, ms-denileukin diftotox, is-denileukin diftotox-VLM, ms-denileukin diftotox-VLM) produced by the methods of the present invention will perform similarly, or better, than Ontak® that is commercially available with regard to clinical efficacies of treating or preventing disease.

Treatment for Tuberculosis.

[0104] As illustrated in FIG. 10, inventors of the present invention believe diphtheria toxin fusion proteins of the present invention will be active against tuberculosis. Denileukin diftotox is known to deplete IL-2-receptor (CD25+)-bearing cells including T regulatory (T_{regs}) cells. T_{regs} cells

express CD25 as well as FoxP3 and are immunosuppressive by their inhibition of Teffector (T_{eff}) cells. T_{eff} cells such as CD4+Thelper (T_h) cells and CD8+cytotoxic T lymphocytes (CTLs) are needed within a tuberculosis granuloma to contain the *M. tuberculosis* bacterial infection. During tuberculosis infection, cellular lesions called granulomas form to contain the infection but are unable to fully eradicate the bacilli. Regulatory T cells (Tregs) are recruited to granulomas, leading to suppression of effector T cell function, potentially contributing to a permissive environment for *M. tuberculosis* persistence and growth. The diphtheria toxin fusion proteins of the present invention are used to deplete Tregs, which express IL-2 receptor, in order to ameliorate immune suppression by these cells during TB infection. FIG. 11 illustrates diphtheria fusion proteins used in the in vivo treatment of subjects (mice) with *M. tuberculosis*. Mice were infected with *M. tb.* strain H37R_v by aerosol infection giving an initial implantation of ~2.8 log₁₀ CFU counts in lungs on day 0. The groups of mice were treated with 750 ng of c-Ontak® intraperitoneally (IP) or intravenously (IV) as one treatment cycle (1×, dosed at week 2 post-infection) or two treatment cycles (2×, dosed at ~day 3 pre-infection and week 2 post-infection). A treatment cycle of denileukin diftotox is defined as 35 mg/kg (750 ng for a typical mouse) given two times, two days apart. RHZ daily treatment by oral gavage was started at week 2. R is rifampin and was given to mice at 10 mg/kg. H is isoniazid and was given to mice at 10 mg/kg. Z is pyrazinamide and was given to mice at 150 mg/kg. The outcome of this study is illustrated in FIGS. 12 and 13.

Treatment for Cancer

[0105] Tregs have also been shown to inhibit anti-tumor immunity, and the cellular expansion of Tregs in tumors generally correlates with poor prognosis in patients. Denileukin diftotox treatment in melanoma patients resulted in transient depletion of Tregs and increased 1 year median overall survival, s-denileukm diftotox and s-denileukin diftotox-VLM of the present invention will be used to deplete Tregs in patients with tumors heavily infiltrated with Tregs as a cancer immunotherapy.

Nucleic Acid and Protein Sequences of s-Ontak-His₆ (“His₆” Disclosed as SEQ ID NO: 23)

[0106] Protein Sequence of *C. diphtherias* derived s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) (theoretical MW 58339) (IL2 portion in Boldface) (SEQ ID NO: 58):

GADDVVDSSKSFVMENFSSYHGTKP

GYVDSIQKGIQPKSGTQGNYYDDWKGFYSTDNKYDAAGYSVDNENPLSG

KAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTTEFIK

RFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRGQ

DAMYEYMAQACAGNRVRRSVGSSSLSCINLDWDVIRDKTCTKIESLKEHGP

IKNKMSSEPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAG

ANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNT

EEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSY

NRPAYSPGHKTHAPTSSSTKKTQLQLEHLLLDLQMLNGINNYKNPKLTR

MLTFKFYMPKKATELKHLCLEELKPLEEVLNLAQSKNFHLRPRDLISN

-continued

INVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLTHHHHH

H

[0107] DNA sequence for *C. diphtheriae*-derived s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) (SEQ ID NO: 59). Alterations to promoter/operator are starred with bold-face. The underlined portion encodes the signal sequence. The first codon of mature s-Ontak-His₆ (“His₆” disclosed as SEQ ID NO: 23) begins at base 149 in larger font and italicized. The codons for the C-terminal His₆ (“His₆” disclosed as SEQ ID NO: 23) begin at base 1709 in larger font and italicized.

1 TTGATTTTCAGAGCACCCCTTATAATTAGGATAGCT**AA**GTCCATTA
TTTTAT

51 GAGTCCTGGTAAGGGGATACGTTGTGAGCAGAAA**ACTGTTTGCG**
TCAATC

101 **TTAATAGGGGCGCTACTGGGGATAGGGGCCCCACCTTCAGCCCA**
TGCAGG

151 CGCTGATGATGTTGTTGATTCTTCTAAATCTTTTGTGATGGAAA
ACTTTT

201 CTTCTGACCACGGGACTAAACCTGGTTATGTAGATTCCATTCAA
AAAGGT

251 ATACAAAAGCCAAAATCTGGTACACAAGGAAATTATGACGATGA
TTGGAA

301 AGGGTTTTATAGTACCGACAATAAATACGACGCTGCGGGATACT
CTGTAG

351 ATAATGAAAACCCGCTCTCTGGAAAAGCTGGAGGCGTGGTCAAA
GTGACG

401 TATCCAGGACTGACGAAGGTTCTCGCACTAAAAGTGGATAATGC
CGAAAC

451 TATTAAGAAAGAGTTAGGTTTAAGTCTCACTGAACCGTTGATGG
AGCAAG

501 TCGGAACGGAAGAGTTTATCAAAAAGGTTCGGTGATGGTGCTTCG
CGTGTA

551 GTGCTCAGCCTTCCCTTCGCTGAGGGGAGTTCTAGCGTTGAATA
TATTAA

601 TAACTGGGAACAGGCGAAAGCGTTAAGCGTAGAAGTTGAGATTA
ATTTTG

651 AAACCCGTGGAAAACGTGGCCAAGATGCGATGTATGAGTATATG
GCTCAA

701 GCCTGTGCAGGAAATCGTGTGAGGCGATCAGTAGGTAGCTCATT
GTCATG

751 CATCAACCTGGATTGGGATGTTATCCGTGATAAACTAAACTA
AGATCG

801 AATCTCTGAAAGAACACGGTCCGATCAAAAACAAAATGAGCGAA
AGCCCG

851 AACAAAAGTGTATCTGAAGAAAAGCTAAACAGTACCTGGAAGA
ATTCCA

901 CCAGACTGCACTGGAACACCCGGAAGTGTCTGAACCTAAGACCG
TTACTG

- continued

951 GTACCAACCCGGTATTCGCTGGTGCTAACTACGCTGCTTGGGCA
GTAAAC

1001 GTTGCTCAGGTTATCGATAGCGAACTGCTGATAACCTGGAAAA
AACTAC

1051 CGCGGCTCTGTCTATCCTGCCGGGTATCGGTAGCGTAATGGGCA
TCGCAG

1101 ACGGCGCCGTTCAACCACAACACTGAAGAAATCGTTGCACAGTCT
ATCGCT

1151 CTGAGCTCTCTGATGGTTGCTCAGGCCATCCCGCTGGTAGGTGA
ACTGGT

1201 TGATATCGGTTTCGCTGCATACAACTTCGTTGAAAGCATCATCA
ACCTGT

1251 TCCAGGTTGTTCACAACTCTTACAACCGCCCGCTTACTCTCCG
GGTCAC

1301 AAGACGCATGCACCTACTTCTAGCTCTACCAAGAAAACCCAGCT
GCAGCT

1351 CGAGCACCTGCTGCTGGATTTCAGATGATCCTGAACGGTATCA
ACAATT

1401 ACAAGAACCCGAACTGACGCGTATGCTGACCTTCAAGTTCTAC
ATGCCG

1451 AAGAAGGCCACCGAACTGAAACACCTGCAGTGTCTAGAAGAAGA
ACTGAA

1501 ACCGCTGGAGGAAGTTCTGAACCTGGCTCAGTCTAAAACTTCC
ACCTGC

1551 GGCCGCGTGACCTGATCTCTAACATCAACGTAATCGTTCTGGAA
CTGAAG

1601 GGCTCTGAAACCACCTTCATGTGTGAATACGCTGATGAGACCGC
AACCAT

1651 CGTAGAATTCTGAACCGTTGGATCACCTTCTGTCTAGTCTATCA
TCTCTA

1701 CCCTGACCCACCATCACCATCATCACTGA < 1711

[0108] All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

[0109] The use of the terms “a” and “an” and “the” and similar referents in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms “comprising,” “having,” “including,” and “containing” are to be construed as open-ended terms (i.e., meaning “including, but not limited to,”) unless otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., “such as”) provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the

specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[0110] Preferred embodiments of this invention are described herein, including the best mode for carrying out the invention. Variations of those preferred embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the described elements of the invention in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

[0111] Embodiments of the disclosure concern methods and/or compositions for treating and/or preventing disorders such as cancer and tuberculosis in which a subject is administered a composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof

[0112] An individual known to having disease such as cancer and/or tuberculosis, suspected of having such a disease, or at risk for having such a disease may be provided an effective amount of a composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof. Those at risk for cancer or tuberculosis may be those individuals having one or more genetic factors, may be of advancing age, and/or may have a family history, for example.

[0113] In particular embodiments of the disclosure, an individual is given an agent for cancer and/or tuberculosis therapy in addition to a composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof. Such additional therapy may include chemotherapy or antimicrobial agents, for example. When combination therapy is employed with a composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, the additional therapy may be given prior to, at the same time as, and/or subsequent to a composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof.

Pharmaceutical Preparations

[0114] Pharmaceutical compositions of the present invention comprise an effective amount of one or more composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, dissolved or dispersed in a pharmaceutically acceptable carrier. The phrases “pharmaceutical or pharmacologically acceptable” refers to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal, such as, for example, a human, as appropriate. The preparation of a pharmaceutical composition that comprises at least one composition of the present invention comprising a nucleic acid or protein sequence such as any one of SEQ ID NOs: 11-15, or fusion proteins thereof, or additional

active ingredient will be known to those of skill in the art in light of the present disclosure, as exemplified by Remington: The Science and Practice of Pharmacy, 21st Ed. Lippincott Williams and Wilkins, 2005, incorporated herein by reference. Moreover, for animal (e.g., human) administration, it will be understood that preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biological Standards.

[0115] As used herein, “pharmaceutically acceptable carrier” includes any and all solvents, dispersion media, coatings, surfactants, antioxidants, preservatives (e.g., antibacterial agents, antifungal agents), isotonic agents, absorption delaying agents, salts, preservatives, drugs, drug stabilizers, gels, binders, excipients, disintegration agents, lubricants, sweetening agents, flavoring agents, dyes, such like materials and combinations thereof, as would be known to one of ordinary skill in the art (see, for example, Remington’s Pharmaceutical Sciences, 18th Ed. Mack Printing Company, 1990, pp. 1289-1329, incorporated herein by reference). Except insofar as any conventional carrier is incompatible with the active ingredient, its use in the pharmaceutical compositions is contemplated.

[0116] The one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, may comprise different types of carriers depending on whether it is to be administered in solid, liquid or aerosol form, and whether it need to be sterile for such routes of administration as injection. The present compositions can be administered intravenously, intradermally, transdermally, intrathecally, intraarterially, intraperitoneally, intranasally, intravaginally, intrarectally, topically, intramuscularly, subcutaneously, mucosally, orally, topically, locally, inhalation (e.g., aerosol inhalation), injection, infusion, continuous infusion, localized perfusion bathing target cells directly, via a catheter, via a lavage, in cremes, in lipid compositions (e.g., liposomes), or by other method or any combination of the foregoing as would be known to one of ordinary skill in the art (see, for example, Remington’s Pharmaceutical Sciences, 18th Ed. Mack Printing Company, 1990, incorporated herein by reference).

[0117] The one or more compositions of the present invention comprising a nucleic acid or protein sequence such as any one of SEQ ID NOs: 11-15, or fusion proteins thereof, may be formulated into a composition in a free base, neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts, e.g., those formed with the free amino groups of a proteinaceous composition, or which are formed with inorganic acids such as for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric or mandelic acid. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as for example, sodium, potassium, ammonium, calcium or ferric hydroxides; or such organic bases as isopropylamine, trimethylamine, histidine or procaine. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as formulated for parenteral administrations such as injectable solutions, or aerosols for delivery to the lungs, or formulated for alimentary administrations such as drug release capsules and the like.

[0118] Further in accordance with the present disclosure, the composition of the present invention suitable for admin-

istration is provided in a pharmaceutically acceptable carrier with or without an inert diluent. The carrier should be assimilable and includes liquid, semi-solid, i.e., pastes, or solid carriers. Except insofar as any conventional media, agent, diluent or carrier is detrimental to the recipient or to the therapeutic effectiveness of a composition contained therein, its use in administrable composition for use in practicing the methods of the present invention is appropriate. Examples of carriers or diluents include fats, oils, water, saline solutions, lipids, liposomes, resins, binders, fillers and the like, or combinations thereof. The composition may also comprise various antioxidants to retard oxidation of one or more component. Additionally, the prevention of the action of microorganisms can be brought about by preservatives such as various antibacterial and antifungal agents, including but not limited to parabens (e.g., methylparabens, propylparabens), chlorobutanol, phenol, sorbic acid, thimerosal or combinations thereof.

[0119] In accordance with the present invention, the composition is combined with the carrier in any convenient and practical manner, i.e., by solution, suspension, emulsification, admixture, encapsulation, absorption and the like. Such procedures are routine for those skilled in the art.

[0120] In a specific embodiment of the present invention, the composition is combined or mixed thoroughly with a semi-solid or solid earner. The mixing can be earned out in any convenient manner such as grinding. Stabilizing agents can be also added in the mixing process in order to protect the composition from loss of therapeutic activity, i.e., denaturation in the stomach. Examples of stabilizers for use in an the composition include buffers, amino acids such as glycine and lysine, carbohydrates such as dextrose, mannose, galactose, fructose, lactose, sucrose, maltose, sorbitol, mannitol, etc.

[0121] In further embodiments, the present invention may concern the use of a pharmaceutical lipid vehicle composition that includes one or more composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, one or more lipids, and an aqueous solvent. As used herein, the term “lipid” will be defined to include any of a broad range of substances that is characteristically insoluble in water and extractable with an organic solvent. This broad class of compounds are well known to those of skill in the art, and as the term “lipid” is used herein, it is not limited to any particular structure. Examples include compounds which contain long-chain aliphatic hydrocarbons and their derivatives. A lipid may be naturally occurring or synthetic (i.e., designed or produced by man). However, a lipid is usually a biological substance. Biological lipids are well known in the art, and include for example, neutral fats, phospholipids, phosphoglycerides, steroids, terpenes, lysolipids, glycosphingolipids, glycolipids, sulphatides, lipids with ether and ester-linked fatty acids and polymerizable lipids, and combinations thereof. Of course, compounds other than those specifically described herein that are understood by one of skill in the art as lipids are also encompassed by the compositions and methods of the present invention.

[0122] One of ordinary skill in the art would be familiar with the range of techniques that can be employed for dispersing a composition in a lipid vehicle. For example, the one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof may be

dispersed in a solution containing a lipid, dissolved with a lipid, emulsified with a lipid, mixed with a lipid, combined with a lipid, covalently bonded to a lipid, contained as a suspension in a lipid, contained or complexed with a micelle or liposome, or otherwise associated with a lipid or lipid structure by any means known to those of ordinary skill in the art. The dispersion may or may not result in the formation of liposomes.

[0123] The actual dosage amount of a composition of the present invention administered to an animal patient can be determined by physical and physiological factors such as bodyweight, severity of condition, the type of disease being treated, previous or concurrent therapeutic interventions, idiopathy of the patient and on the route of administration. Depending upon the dosage and the route of administration, the number of administrations of a preferred dosage and/or an effective amount may vary according to the response of the subject. The practitioner responsible for administration will, in any event, determine the concentration of active ingredient(s) in a composition and appropriate dose(s) for the individual subject.

[0124] In certain embodiments, pharmaceutical compositions may comprise, for example, at least about 0.1% of an active compound. In other embodiments, an active compound may comprise between about 2% to about 75% of the weight of the unit, or between about 25% to about 60%, for example, and any range derivable therein. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared in such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

[0125] In other non-limiting examples, a dose may also comprise from about 1 microgram/kg/body weight, about 5 microgram/kg/body weight, about 10 microgram/kg/body weight, about 50 microgram/kg/body weight, about 100 microgram/kg/body weight, about 200 microgram/kg/body weight, about 350 microgram/kg/body weight, about 500 microgram/kg/body weight, about 1 milligram/kg/body weight, about 5 milligram/kg/body weight, about 10 milligram/kg/body weight, about 50 milligram/kg/body weight, about 100 milligram/kg/body weight, about 200 milligram/kg/body weight, about 350 milligram/kg/body weight, about 500 milligram/kg/body weight, to about 1000 mg/kg/body weight or more per administration, and any range derivable therein. In non-limiting examples of a derivable range from the numbers listed herein, a range of about 5 mg/kg/body weight to about 100 mg/kg/body weight, about 5 microgram/kg/body weight to about 500 milligram/kg/body weight, etc., can be administered, based on the numbers described above.

[0126] In further embodiments, a pharmaceutical composition of the invention as described in any of the previous embodiments comprises greater than about 80% purity of a polypeptide of the invention. In other embodiments, the pharmaceutical composition comprises greater than about 81%, greater than about 82%, greater than about 83%, greater than about 84%, greater than about 85%, greater than about 86%, greater than about 87%, greater than about 88%, greater than about 89%, greater than about 90%, greater than

about 91%, greater than about 92%, greater than about 93%, greater than about 94%, greater than about 95%, greater than about 96%, greater than about 97%, greater than about 98%, greater than about 99%, or about 100% purity of a polypeptide of the invention. In other embodiments, the pharmaceutical composition comprises about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or about 100% purity of a polypeptide of the invention. In other embodiments, the pharmaceutical composition comprises from about 80% to about 100%, from about 80% to about 97%, from about 80% to about 95%, from about 80% to about 90%, from about 80% to about 85%, from about 85% to about 100%, from about 85% to about 97%, from about 85% to about 95%, from about 85% to about 90%, from about 90% to about 100%, from about 90% to about 97%, from about 90% to about 95%, from about 95% to about 100%, or from about 95% to about 97% purity of a polypeptide of the invention, or any other range thereof.

[0127] In further embodiments, a pharmaceutical composition of the invention as described in any of the previous embodiments comprises greater than about 80% aggregate-free, full-length, monomeric polypeptide of the invention. In other embodiments, the pharmaceutical composition comprises greater than about 81%, greater than about 82%, greater than about 83%, greater than about 84%, greater than about 85%, greater than about 86%, greater than about 87%, greater than about 88%, greater than about 89%, greater than about 90%, greater than about 91%, greater than about 92%, greater than about 93%, greater than about 94%, greater than about 95%, greater than about 95%, greater than about 96%, greater than about 97%, greater than about 98%, greater than about 99%, or about 100% aggregate-free, full-length, monomeric polypeptide of the invention. In other embodiments, the pharmaceutical composition comprises about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or about 100% aggregate-free, full-length, monomeric polypeptide of the invention. In other embodiments, the pharmaceutical composition comprises from about 80% to about 100%, from about 80% to about 97%, from about 80% to about 95%, from about 80% to about 90%, from about 80% to about 85%, from about 85% to about 100%, from about 85% to about 97%, from about 85% to about 95%, from about 85% to about 90%, from about 90% to about 100%, from about 90% to about 97%, from about 90% to about 95%, from about 95% to about 100%, or from about 95% to about 97% aggregate-free, full-length, monomeric polypeptide of the invention, or any other range thereof.

[0128] In further embodiments, a pharmaceutical composition of the invention comprises greater than about 80% purity of a polypeptide of the invention (or any other range or amount described herein) and greater than about 80% aggregate-free, full-length, monomeric polypeptide of the invention (or any other range or amount described herein).

[0129] In further embodiments, a polypeptide of such pharmaceutical compositions comprises a histidine (His) tag. In some embodiments, the His tag has six or nine His residues. In other embodiments, the His tag is at the C-ter-

minus of the polypeptide. In other embodiments, a polypeptide of such pharmaceutical compositions does not comprise a His tag.

Alimentary Compositions and Formulations

[0130] In one embodiment of the present disclosure, the one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, are formulated to be administered via an alimentary route. Alimentary routes include all possible routes of administration in which the composition is in direct contact with the alimentary tract. Specifically, the pharmaceutical compositions disclosed herein may be administered orally, buccally, rectally, or sublingually. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

[0131] In certain embodiments, the active compounds may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (Mathiowitz et al., 1997; Hwang et al., 1998; U.S. Pat. Nos. 5,641,515; 5,580,579 and 5,792,451, each specifically incorporated herein by reference in its entirety). The tablets, troches, pills, capsules and the like may also contain the following: a binder, such as, for example, gum tragacanth, acacia, cornstarch, gelatin or combinations thereof; an excipient, such as, for example, dicalcium phosphate, mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate or combinations thereof; a disintegrating agent, such as, for example, corn starch, potato starch, alginic acid or combinations thereof; a lubricant, such as, for example, magnesium stearate; a sweetening agent, such as, for example, sucrose, lactose, saccharin or combinations thereof; a flavoring agent, such as, for example peppermint, oil of Wintergreen, cherry flavoring, orange flavoring, etc. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. When the dosage form is a capsule, it may contain, in addition to materials of the above type, carriers such as a liquid carrier. Gelatin capsules, tablets, or pills may be enterically coated. Enteric coatings prevent denaturation of the composition in the stomach or upper bowel where the pH is acidic. See, e.g., U.S. Pat. No. 5,629,001. Upon reaching the small intestines, the basic pH therein dissolves the coating and permits the composition to be released and absorbed by specialized cells, e.g., epithelial enterocytes and Peyer's patch M cells. A syrup or elixir may contain the active compound sucrose as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

[0132] For oral administration the compositions of the present disclosure may alternatively be incorporated with one or more excipients in the form of a mouthwash, denti-

frice, buccal tablet, oral spray, or sublingual orally-administered formulation. For example, a mouthwash may be prepared incorporating the active ingredient in the required amount in an appropriate solvent, such as a sodium borate solution (Dobell's Solution). Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

[0133] Additional formulations which are suitable for other modes of alimentary administration include suppositories. Suppositories are solid dosage forms of various weights and shapes, usually medicated, for insertion into the rectum. After insertion, suppositories soften, melt or dissolve in the cavity fluids. In general, for suppositories, traditional carriers may include, for example, polyalkylene glycols, triglycerides or combinations thereof. In certain embodiments, suppositories may be formed from mixtures containing, for example, the active ingredient in the range of about 0.5% to about 10%, and preferably about 1% to about 2%.

Parenteral Compositions and Formulations

[0134] In further embodiments, one or more composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, may be administered via a parenteral route. As used herein, the term "parenteral" includes routes that bypass the alimentary tract. Specifically, the pharmaceutical compositions disclosed herein may be administered for example, but not limited to intravenously, intradermally, intramuscularly, intraarterially, intrathecally, subcutaneous, or intraperitoneally U.S. Pat. Nos. 6,7537,514; 6,613,308; 5,466,468; 5,543,158; 5,641,515; and 5,399,363 (each specifically incorporated herein by reference in its entirety).

[0135] Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms. The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U.S. Pat. 5,466,468, specifically incorporated herein by reference in its entirety). In all cases the form must be sterile and must be fluid to the extent that easy injectability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (i.e., glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by

various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

[0136] For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous, and intraperitoneal administration. In this connection, sterile aqueous media that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in isotonic NaCl solution and either added hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

[0137] Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof. A powdered composition is combined with a liquid carrier such as, e.g., water or a saline solution, with or without a stabilizing agent.

Miscellaneous Pharmaceutical Compositions and Formulations

[0138] In other preferred embodiments of the invention, the one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, may be formulated for administration via various miscellaneous routes, for example, topical (i.e., transdermal) administration, mucosal administration (intranasal, vaginal, etc.) and/or inhalation.

[0139] Pharmaceutical compositions for topical administration may include the active compound formulated for a medicated application such as an ointment, paste, cream or powder. Ointments include all oleaginous, adsorption, emulsion and water-soluble based compositions for topical application, while creams and lotions are those compositions that include an emulsion base only. Topically administered medications may contain a penetration enhancer to facilitate adsorption of the active ingredients through the skin. Suitable

penetration enhancers include glycerin, alcohols, alkyl methyl sulfoxides, pyrrolidones and laurocapram. Possible bases for compositions for topical application include polyethylene glycol, lanolin, cold cream and petrolatum as well as any other suitable absorption, emulsion or water-soluble ointment base. Topical preparations may also include emulsifiers, gelling agents, and antimicrobial preservatives as necessary to preserve the active ingredient and provide for a homogenous mixture. Transdermal administration of the present invention may also comprise the use of a "patch". For example, the patch may supply one or more active substances at a predetermined rate and in a continuous manner over a fixed period of time.

[0140] In certain embodiments, the pharmaceutical compositions may be delivered by eye drops, intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering compositions directly to the lungs via nasal aerosol sprays has been described e.g., in U.S. Pat. Nos. 5,756,353 and 5,804,212 (each specifically incorporated herein by reference in its entirety). Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga et al., 1998) and lysophosphatidyl-glycerol compounds (U.S. Pat. No. 5,725,871, specifically incorporated herein by reference in its entirety) are also well-known in the pharmaceutical arts. Likewise, transmucosal drug delivery in the form of a polytetrafluoroethylene support matrix is described in U.S. Pat. No. 5,780,045 (specifically incorporated herein by reference in its entirety).

[0141] The term aerosol refers to a colloidal system of finely divided solid or liquid particles dispersed in a liquefied or pressurized gas propellant. The typical aerosol of the present invention for inhalation will consist of a suspension of active ingredients in liquid propellant or a mixture of liquid propellant and a suitable solvent. Suitable propellants include hydrocarbons and hydrocarbon ethers. Suitable containers will vary according to the pressure requirements of the propellant. Administration of the aerosol will vary according to subject's age, weight and the severity and response of the symptoms.

Kits of the Disclosure

[0142] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, one or more composition of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, may be comprised in a kit.

[0143] The kits may comprise a suitably aliquoted of one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, and in some cases, one or more additional agents. The component(s) of the kits may be packaged either in aqueous media or in lyophilized form. The container means of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which a component may be placed, and preferably, suitably aliquoted. Where there are more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a vial. The kits of the present invention also will typically include a means for containing the one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs:

11-15, or fusion proteins thereof, and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

[0144] When the components of the kit are provided in one and/or more liquid solutions, the liquid solution is an aqueous solution, with a sterile aqueous solution being particularly preferred. The one or more compositions of the present invention comprising a nucleic acid or protein sequence such as anyone of SEQ ID NOs: 11-15, or fusion proteins thereof, may be formulated into a syringeable composition. In which case, the container means may itself be a syringe, pipette, and/or other such like apparatus, from which the formulation may be applied to an infected area of the body, injected into an animal, and/or even applied to and/or mixed with the other components of the kit.

[0145] However, the components of the kit may be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent. It is envisioned that the solvent may also be provided in another container means.

[0146] Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

SEQUENCE LISTING		
<160> NUMBER OF SEQ ID NOS: 59		
<210> SEQ ID NO 1		
<211> LENGTH: 19		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide		
<400> SEQUENCE: 1		
ttaggatagc taagtccat	19	
<210> SEQ ID NO 2		
<211> LENGTH: 42		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide		
<400> SEQUENCE: 2		
ttgatttcag agcaccctta taattaggat agctaagtc at	42	
<210> SEQ ID NO 3		
<211> LENGTH: 1711		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide		
<400> SEQUENCE: 3		
ttgatttcag agcaccctta taattaggat agctaagtc attattttat gagtectggt	60	
aaggggatac gttgtgagca gaaaactgtt tgcgtcaatc ttaatagggg cgctactggg	120	
gatagggggc ccaccttcag cccatgcagg cgctgatgat gttgttgatt cttctaaatc	180	
ttttgtgatg gaaaactttt cttcgtacca cgggactaaa cctggttatg tagattccat	240	
tcaaaaaggt atacaaaagc caaaatctgg tacacaagga aattatgacg atgattggaa	300	
agggttttat agtaccgaca ataaatacga cgctgcggga tactctgtag ataatgaaaa	360	
cccgtctctt ggaaaagctg gaggcgtggt caaagtgacg tatccaggac tgacgaaggt	420	
tctcgcaact aaagtggata atgccgaaac tattaagaaa gagttagggt taagtctcac	480	
tgaaccgttg atggagcaag tcggaacgga agagtttatc aaaagggtcg gtgatggtgc	540	
ttcgcgtgta gtgctcagcc ttcccttcgc tgaggggagt tctagcgttg aatatattaa	600	

-continued

taactgggaa caggcgaaag cgtaaagcgt agaacttgag attaatTTtg aaaccctggg	660
aaaacgtggc caagatgca tgtatgagta tatggctcaa gcctgtgcag gaaatcgtgt	720
caggcgatca gtaggtagct cattgtcatg catcaacctg gattgggatg ttatccgtga	780
taaaactaaa actaagatcg aatctctgaa agaacacggg ccgatcaaaa acaaaatgag	840
cgaaagcccg aacaaaactg tatctgaaga aaaagctaaa cagtacctgg aagaattcca	900
ccagactgca ctggaacacc cggaactgtc tgaacttaag accgttactg gtaccaaccc	960
ggtattcgct ggtgctaact acgtgcttg ggagtaaac gttgctcagg ttatcgatag	1020
cgaaactgct gataacctgg aaaaaactac cgcggctctg tctatcctgc cgggtatcgg	1080
tagcgtaatg ggcatcgag acggcgccgt tcaccacaac actgaagaaa tcgttgacac	1140
gtctatcgct ctgagctctc tgatgggtgc tcaggccatc ccgctggtag gtgaactggg	1200
tgatatcggt ttcgctgcat acaacttctg tgaagcatc atcaacctgt tccagggtgt	1260
tcacaactct tacaaccgcc cggttactc tccgggtcac aagacgcatg cacctacttc	1320
tagctctacc aagaaaaccc agctgcagct cgagcacctg ctgctggatt tgcagatgat	1380
cctgaacggg atcaacaatt acaagaaccc gaaactgacg cgtatgctga ccttcaagtt	1440
ctacatgccg aagaaggcca ccgaactgaa acacctgcag tgtctagaag aagaactgaa	1500
accgctggag gaagttctga acctggctca gtctaaaaac ttccacctgc ggccgctga	1560
cctgatctct aacatcaacg taatcgttct ggaactgaag ggctctgaaa ccaccttcac	1620
gtgtgaatac gctgatgaga ccgaaccat cgtagaatc ctgaaccgtt ggatcacctt	1680
ctgtcagtct atcatctcta cctgacctg a	1711
<210> SEQ ID NO 4	
<211> LENGTH: 75	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide	
<400> SEQUENCE: 4	
gtgagcagaa aactgTTTgc gtcaatctta atagggggcg tactggggat aggggccccca	60
ccttcagccc atgca	75
<210> SEQ ID NO 5	
<211> LENGTH: 25	
<212> TYPE: PRT	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide	
<400> SEQUENCE: 5	
Met Ser Arg Lys Leu Phe Ala Ser Ile Leu Ile Gly Ala Leu Leu Gly	
1 5 10 15	
Ile Gly Ala Pro Pro Ser Ala His Ala	
20 25	
<210> SEQ ID NO 6	
<211> LENGTH: 1566	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	

-continued

<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide		
<400> SEQUENCE: 6		
atgggcgctg atgatgttgt tgattcttct aaatcttttg tgatggaaaa cttttcttcg	60	
taccacggga ctaaacctgg ttatgtagat tccattcaaa aaggatatac aaagccaaaa	120	
tctggtacac aaggaaatta tgacgatgat tggaaagggt tttatagtag cgacaataaa	180	
tacgacgctg cgggatactc tgtagataat gaaaaccgct tctctggaaa agctggaggc	240	
gtggtcaaag tgacgtatcc aggactgacg aaggttctcg cactaaaagt ggataatgcc	300	
gaaactatta agaaagagtt aggtttaagt ctactgaac cgttgatgga gcaagtcgga	360	
acggaagagt ttatcaaaaag gttcgggtgat ggtgcttcgc gtgtagtgct cagccttccc	420	
ttcgctgagg ggagttctag cggtgaatat attaataact gggaacaggc gaaagcgta	480	
agcgtagaac ttgagattaa ttttgaaacc cgtggaaaac gtggccaaga tgcgatgtat	540	
gagtatatgg ctcaagcctg tgcaggaaat cgtgtcaggc gatcagtagg tagctcattg	600	
tcatgcatca acctggattg ggatgttatc cgtgataaaa ctaaaactaa gatcgaatct	660	
ctgaaagaac acggtccgat caaaaacaaa atgagcgaaa gccgaacaa aactgtatct	720	
gaagaaaaag ctaaacagta cctggaagaa ttccaccaga ctgcactgga acaccggaa	780	
ctgtctgaac ttaagaccgt tactggtacc aacccggtat tcgctggtgc taactacgct	840	
gcttgggcag taaacgttgc tcaggttatc gatagcgaaa ctgctgataa cctggaaaaa	900	
actaccgagg ctctgtctat cctgccgggt atcggtagcg taatgggcat cgcagacggc	960	
gccgttcacc acaacactga agaaatcggt gcacagtcta tcgctctgag ctctctgatg	1020	
gttgetcagg ccatcccgct ggtaggtgaa ctggttgata tcggtttcgc tgcatacaac	1080	
ttcgttgaaa gcatcatcaa cctgttccag gttgttcaca actcttaca ccgccggct	1140	
tactctcggg gtcacaagac gcatgcacct acttctagct ctaccaagaa aaccagctg	1200	
cagctcgagc acctgctgct ggatttgtag atgacctga acggtatcaa caattacaag	1260	
aaccggaaac tgacgcgtat gctgacctc aagtctaca tgccgaagaa ggccaccgaa	1320	
ctgaaacacc tgcagtgtct agaagaagaa ctgaaaccgc tggaggaagt tctgaacctg	1380	
gctcagtcta aaaacttcca cctgcggccg cgtgacctga tctctaakat caacgtaatc	1440	
gttctggaac tgaagggtc tgaaccacc ttcatgtgtg aatacgctga tgagaccgca	1500	
accatcgtag aattcctgaa ccgttggtac accttctgtc agtctatcat ctctaccctg	1560	
acctga	1566	
<210> SEQ ID NO 7		
<211> LENGTH: 1566		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide		
<400> SEQUENCE: 7		
atgggcgctg atgatgttgc tgattcttct aaatcttttg tgatggaaaa cttttcttcg	60	
taccacggga ctaaacctgg ttatgtagat tccattcaaa aaggatatac aaagccaaaa	120	
tctggtacac aaggaaatta tgacgatgat tggaaagggt tttatagtag cgacaataaa	180	

-continued

tacgacgctg cgggatactc tgtagataat gaaaacccgc tctctggaaa agctggaggc	240
gtggtcaaag tgacgtatcc aggactgacg aaggttctcg cactaaaagt ggataatgcc	300
gaaactatta agaaagagtt aggtttaagt ctcaactgaac cgttgatgga gcaagtcgga	360
acggaagagt ttatcaaaaag gttcggtgat ggtgcttcgc gtgtagtgct cagccttccc	420
ttcgtgaggg ggagttctag cgttgaatat attaataact gggaacaggc gaaagcgta	480
agcgtagaac ttgagattaa ttttgaaacc cgtggaaaac gtggccaaga tgcgatgtat	540
gagtatatgg ctcaagcctg tgcaggaaat cgtgtcaggc gatcagtagg tagctcattg	600
tcatgcatca acctggattg ggatgttatc cgtgataaaa ctaaaactaa gatcgaatct	660
ctgaaagaac acggtccgat caaaaacaaa atgagcgaaa gcccgaaaca aactgtatct	720
gaagaaaaag ctaaacagta cctggaagaa ttccaccaga ctgcactgga acacccgga	780
ctgtctgaac ttaagaccgt tactggtacc aacccggtat tcgctgggtc taactacgt	840
gcttgggcag taaacgttgc tcaggttatc gatagcgaaa ctgctgataa cctggaaaaa	900
actaccgcgg ctctgtctat cctgccgggt atcggtagcg taatgggcat cgcagacggc	960
gccgttcacc acaacactga agaaatcggt gcacagtcta tcgctctgag ctctctgatg	1020
gttgctcagg ccatcccgct ggtaggtgaa ctggttgata tcggtttcgc tgcatacaac	1080
ttcgttgaaa gcatcatcaa cctgttccag gttgttcaca actcttaca cgcgccggct	1140
tactctccgg gtcacaagac gcatgcacct acttctagct ctaccaagaa aaccagctg	1200
cagctcgagc acctgctgct ggatttgag atgatacctga acggtatcaa caattacaag	1260
aaccgaaaac tgacgcgtat gctgaccttc aagtctaca tgccgaagaa ggccaccgaa	1320
ctgaaacacc tgctgcagtg tctagaagaa gaactgaaac cgctggagga agttctgaac	1380
ctggctcagt ctaaaaactt ccacctgcgg ccgctgacc tgatctctaa catcaacgta	1440
atcgttctgg aactgaaggg ctctgaaacc accttcattg gtgaatacgc tgatgagacc	1500
gcaaccatcg tagaattcct gaaccgttgg atcaccttct gtcagtctat catctctacc	1560
ctgacc	1566
<210> SEQ ID NO 8	
<211> LENGTH: 1638	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide	
<400> SEQUENCE: 8	
gtgagcagaa aactgtttgc gtcaatctta ataggggcgc tactggggat aggggcccc	60
ccttcagccc atgcaggcgc tgatgatgtt gctgattctt ctaaattctt tgtgatggaa	120
aacttttctt cgtaccacgg gactaaacct ggttatgtag attccattca aaaaggtata	180
caaaagccaa aatctggtac acaaggaaat tatgacgatg attggaaagg gttttatagt	240
accgacaata aatacgacgc tgcgggatac tctgtagata atgaaaaccc gctctctgga	300
aaagctggag gcgtggtcaa agtgacgtat ccaggactga cgaaggttct cgcactaaaa	360
gtggataatg ccgaaactat taagaaagag ttaggtttaa gtctcactga accgttgatg	420
gagcaagtcg gaacggaaga gtttatcaaa aggttcggtg atggtgcttc gcgtgtagtg	480

-continued

ctcagccttc ccttcgctga ggggagttct agcgttgaat atattaataa ctgggaacag	540
gcgaaagcgt taagcgtaga acttgagatt aattttgaaa cccgtggaaa acgtggccaa	600
gatgcgatgt atgagtatat ggctcaagcc tgtgcaggaa atcgtgtcag gcgatcagta	660
ggtagctcat tgtcatgcat caacctggat tgggatgtta tccgtgataa aactaaaact	720
aagatcgaat ctctgaaaga acacgggtccg atcaaaaaca aaatgagcga aagcccgaac	780
aaaactgtat ctgaagaaaa agctaaacag tacctggaag aattccacca gactgcactg	840
gaacacccgg aactgtctga acttaagacc gttactggta ccaacccggt attcgctggg	900
gctaactacg ctgcttgggc agtaaacgtt gctcaggtta tcgatagcga aactgctgat	960
aacctggaaa aaactaccgc ggctctgtct atcctgccgg gtatcggtag cgtaatgggc	1020
atcgcagacg gcgccgttca ccacaacact gaagaaatcg ttgcacagtc tatcgctctg	1080
agctctctga tggttgctca ggccatcccg ctggtagggtg aactggttga tatcggtttc	1140
gctgcataca acttcgttga aagcatcatc aacctgttcc aggttggttca caactcttac	1200
aaccgcccgg cttactctcc gggtcacaag acgcatgcac ctacttctag ctctaccaag	1260
aaaaccacgc tgcagctcga gcacctgctg ctggatttgc agatgatcct gaacggtatc	1320
aacaattaca agaacccgaa actgacgcgt atgctgacct tcaagttcta catgccgaag	1380
aaggccaccg aactgaaaca cctgcagtgt ctagaagaag aactgaaacc gctggaggaa	1440
gttctgaacc tggctcagtc taaaaacttc cacctgcggc cgcgtgacct gatctctaac	1500
atcaacgtaa tcgttctgga actgaagggc tctgaaacca ccttcatgtg tgaatacgct	1560
gatgagaccg caaccatcgt agaattcctg aaccgttggg tcaccttctg tcagtctatc	1620
atctctaccc tgacctga	1638
<210> SEQ ID NO 9	
<211> LENGTH: 402	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide	
<400> SEQUENCE: 9	
gcacctactt ctagctctac caagaaaacc cagctgcagc tcgagcacct gctgctggat	60
ttgcagatga tcctgaacgg tatcaacaat tacaagaacc cgaaactgac gcgtatgctg	120
accttcaagt tctacatgcc gaagaaggcc accgaactga aacacctgca gtgtctagaa	180
gaagaactga aaccgctgga ggaagttctg aacctggctc agtctaaaaa cttccacctg	240
cggccgcgtg acctgatctc taacatcaac gtaatcgttc tggaactgaa gggctctgaa	300
accaccttca tgtgtgaata cgctgatgag accgcaacca tcgtagaatt cctgaaccgt	360
tggatcacct tctgtcagtc tatcatctct accctgacct ga	402
<210> SEQ ID NO 10	
<211> LENGTH: 521	
<212> TYPE: PRT	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide	
<400> SEQUENCE: 10	

-continued

Met	Gly	Ala	Asp	Asp	Val	Val	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	
1				5					10					15		
Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	
			20					25					30			
Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	
		35					40					45				
Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	
	50					55					60					
Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	
65					70					75					80	
Val	Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	
				85					90						95	
Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	
			100					105					110			
Glu	Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	
		115					120					125				
Gly	Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	
	130					135					140					
Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	
145					150					155					160	
Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	
				165					170					175		
Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	
			180					185					190			
Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	
		195					200					205				
Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	
	210					215					220					
Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	
225					230					235					240	
Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	
				245					250					255		
Glu	His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	
			260					265					270			
Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	
		275					280					285				
Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	
	290					295					300					
Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	
305					310					315					320	
Ala	Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	
				325					330					335		
Ser	Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	
			340					345					350			
Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	
		355					360					365				
Phe	Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	
	370					375					380					
His	Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	
385					390					395					400	
Gln	Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	

-continued

405																410																415																																			
Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe																																																				
420																425																430																																			
Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu																																																				
435																440																445																																			
Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys																																																				
450																455																460																																			
Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile																																																				
465																470																475																480																			
Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala																																																				
485																490																495																																			
Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe																																																				
500																505																510																																			
Cys	Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr																																																											
515																520																																																			
<210> SEQ ID NO 11																																																																			
<211> LENGTH: 560																																																																			
<212> TYPE: PRT																																																																			
<213> ORGANISM: Corynebacterium diphtheriae																																																																			
<400> SEQUENCE: 11																																																																			
Met	Ser	Arg	Lys	Leu	Phe	Ala	Ser	Ile	Leu	Ile	Gly	Ala	Leu	Leu	Gly																																																				
1																	5																	10																	15																
Ile	Gly	Ala	Pro	Pro	Ser	Ala	His	Ala	Gly	Ala	Asp	Asp	Val	Val	Asp																																																				
20																25																30																																			
Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr																																																				
35																40																45																																			
Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys																																																				
50																55																60																																			
Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser																																																				
65																70																75																80																			
Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn																																																				
85																90																95																																			
Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val	Thr	Tyr	Pro	Gly																																																				
100																105																110																																			
Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys																																																				
115																120																125																																			
Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly																																																				
130																135																140																																			
Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val																																																				
145																150																155																160																			
Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn																																																				
165																170																175																																			
Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe																																																				
180																185																190																																			
Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala																																																				
195																200																205																																			
Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu																																																				
210																215																220																																			
Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr																																																				
225																230																235																240																			

-continued

Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser	
				245					250					255		
Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu	
			260					265					270			
Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu	
		275					280					285				
Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala	
	290					295					300					
Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp	
305					310					315					320	
Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly	
				325					330					335		
Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu	
			340					345					350			
Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala	
		355					360					365				
Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn	
	370					375					380					
Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr	
385					390					395					400	
Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	Gln	Pro	Phe	Leu	His	
				405					410					415		
Asp	Gly	Tyr	Ala	Val	Ser	Trp	Asn	Thr	Val	Glu	Asp	Ser	Ile	Ile	Arg	
			420					425					430			
Thr	Gly	Phe	Gln	Gly	Glu	Ser	Gly	His	Asp	Ile	Lys	Ile	Thr	Ala	Glu	
		435					440					445				
Asn	Thr	Pro	Leu	Pro	Ile	Ala	Gly	Val	Leu	Leu	Pro	Thr	Ile	Pro	Gly	
	450					455					460					
Lys	Leu	Asp	Val	Asn	Lys	Ser	Lys	Thr	His	Ile	Ser	Val	Asn	Gly	Arg	
465					470					475					480	
Lys	Ile	Arg	Met	Arg	Cys	Arg	Ala	Ile	Asp	Gly	Asp	Val	Thr	Phe	Cys	
				485					490					495		
Arg	Pro	Lys	Ser	Pro	Val	Tyr	Val	Gly	Asn	Gly	Val	His	Ala	Asn	Leu	
			500					505					510			
His	Val	Ala	Phe	His	Arg	Ser	Ser	Ser	Glu	Lys	Ile	His	Ser	Asn	Glu	
		515					520					525				
Ile	Ser	Ser	Asp	Ser	Ile	Gly	Val	Leu	Gly	Tyr	Gln	Lys	Thr	Val	Asp	
	530					535					540					
His	Thr	Lys	Val	Asn	Ser	Lys	Leu	Ser	Leu	Phe	Phe	Glu	Ile	Lys	Ser	
545					550					555					560	
<210> SEQ ID NO 12																
<211> LENGTH: 545																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 12																
Met	Ser	Arg	Lys	Leu	Phe	Ala	Ser	Ile	Leu	Ile	Gly	Ala	Leu	Leu	Gly	
1				5					10					15		
Ile	Gly	Ala	Pro	Pro	Ser	Ala	His	Ala	Gly	Ala	Asp	Asp	Val	Val	Asp	
		20						25					30			

-continued

Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr	
		35					40					45				
Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys	
	50					55					60					
Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser	
65					70					75					80	
Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn	
			85						90					95		
Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val	Thr	Tyr	Pro	Gly	
			100					105						110		
Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys	
		115					120					125				
Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly	
	130					135					140					
Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val	
145					150					155					160	
Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn	
				165					170					175		
Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe	
			180					185					190			
Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala	
		195					200					205				
Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu	
	210					215					220					
Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr	
225					230					235					240	
Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser	
				245					250					255		
Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu	
			260					265					270			
Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu	
		275					280					285				
Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala	
	290					295					300					
Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp	
305					310					315					320	
Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly	
				325					330					335		
Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu	
			340					345					350			
Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala	
		355					360					365				
Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn	
	370					375					380					
Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr	
385					390					395					400	
Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	His	Ala	Pro	Thr	Ser	
				405					410					415		
Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	Leu	Glu	His	Leu	Leu	Leu	Asp	
			420					425					430			

-continued

Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu	
		435					440					445				
Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	
		450				455					460					
Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	
465					470					475					480	
Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	
				485					490					495		
Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	
			500					505					510			
Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	
		515					520					525				
Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	Gln	Ser	Ile	Ile	Ser	Thr	Leu	
	530					535					540					
Thr																
545																
<210> SEQ ID NO 13																
<211> LENGTH: 520																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 13																
Gly	Ala	Asp	Asp	Val	Val	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	
1				5					10					15		
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	
			20					25				30				
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	
		35				40						45				
Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	
	50					55					60					
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	
65				70						75					80	
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	
			85						90					95		
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	
			100					105					110			
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	
		115					120					125				
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	
	130					135					140					
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	
145					150					155					160	
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	
			165						170					175		
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	
			180					185					190			
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	
		195				200						205				
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	
	210					215						220				

-continued

Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	
225					230					235					240	
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	
				245					250						255	
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	
			260					265						270		
Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	
		275					280					285				
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	
	290					295					300					
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	
305					310					315					320	
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	
				325					330						335	
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	
			340					345						350		
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	
		355					360						365			
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	
		370					375					380				
Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	
385					390						395				400	
Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	
				405					410						415	
Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	
			420					425						430		
Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	
		435					440					445				
Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	
						455						460				
Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	
465					470					475					480	
Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	
				485					490						495	
Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	
			500					505						510		
Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr									
		515					520									
<210> SEQ ID NO 14																
<211> LENGTH: 545																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 14																
Met	Ser	Arg	Lys	Leu	Phe	Ala	Ser	Ile	Leu	Ile	Gly	Ala	Leu	Leu	Gly	
1				5					10					15		
Ile	Gly	Ala	Pro	Pro	Ser	Ala	His	Ala	Gly	Ala	Asp	Asp	Val	Ala	Asp	
			20					25					30			
Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr	

-continued																	
35						40						45					
Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys		
50						55					60						
Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser		
65					70					75					80		
Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn		
				85					90					95			
Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val	Thr	Tyr	Pro	Gly		
			100					105						110			
Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys		
			115				120					125					
Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly		
	130					135					140						
Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val		
145					150					155					160		
Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn		
				165					170						175		
Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe		
			180					185						190			
Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala		
		195					200						205				
Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu		
	210					215					220						
Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr		
225					230					235					240		
Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser		
				245					250					255			
Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu		
			260					265						270			
Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu		
		275					280						285				
Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala		
						295					300						
Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp		
305					310					315					320		
Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly		
				325					330					335			
Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu		
			340					345					350				
Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala		
			355				360						365				
Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn		
						375					380						
Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr		
385					390					395					400		
Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	His	Ala	Pro	Thr	Ser		
				405					410					415			
Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	Leu	Glu	His	Leu	Leu	Leu	Asp		
			420					425					430				
Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu		
		435					440					445					

-continued

Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	
450						455					460					
Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	
465					470					475					480	
Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	
				485					490					495		
Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	
			500					505					510			
Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	
		515					520					525				
Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	Gln	Ser	Ile	Ile	Ser	Thr	Leu	
530						535					540					
Thr																
545																
<210> SEQ ID NO 15																
<211> LENGTH: 520																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 15																
Gly	Ala	Asp	Asp	Val	Ala	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	
1				5					10					15		
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	
			20					25					30			
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	
		35					40					45				
Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	
	50					55					60					
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	
65					70					75					80	
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	
				85					90					95		
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	
			100					105					110			
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	
		115					120					125				
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	
	130						135				140					
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	
145					150					155					160	
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	
				165					170					175		
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	
			180					185					190			
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	
		195					200					205				
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	
	210					215					220					
Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	

-continued															
225				230					235					240	
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu
				245					250					255	
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val
			260					265					270		
Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val
		275					280					285			
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu
	290					295					300				
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala
305					310					315					320
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser
			325						330					335	
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp
			340					345					350		
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe
	355						360					365			
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His
	370					375					380				
Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln
385					390					395					400
Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn
			405						410					415	
Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr
			420					425					430		
Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu
		435					440					445			
Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn
	450					455					460				
Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val
465					470					475					480
Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp
			485						490					495	
Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys
			500					505					510		
Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr								
		515					520								
<210> SEQ ID NO 16															
<211> LENGTH: 522															
<212> TYPE: PRT															
<213> ORGANISM: Artificial Sequence															
<220> FEATURE:															
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide															
<400> SEQUENCE: 16															
Met	Gly	Ala	Asp	Asp	Val	Ala	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu
1				5					10					15	
Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile
			20					25					30		
Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp
		35					40					45			

-continued

Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	
50						55				60						
Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	
65				70					75					80		
Val	Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	
			85					90						95		
Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	
			100					105					110			
Glu	Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	
		115					120					125				
Gly	Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	
	130					135					140					
Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	
145					150				155						160	
Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	
				165					170					175		
Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	
			180					185					190			
Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	
		195					200					205				
Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	
	210					215					220					
Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	
225					230					235					240	
Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	
				245					250					255		
Glu	His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	
			260					265					270			
Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	
		275					280					285				
Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	
	290					295					300					
Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	
305					310					315					320	
Ala	Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	
				325					330					335		
Ser	Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	
			340					345					350			
Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	
		355					360					365				
Phe	Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	
	370					375					380					
His	Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	
385					390					395					400	
Gln	Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	
			405						410					415		
Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	
			420					425					430			
Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Leu	Gln	Cys	Leu	
		435					440					445				
Glu	Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	

-continued																			
450					455					460									
Lys	Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val				
465					470					475					480				
Ile	Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr				
				485					490					495					
Ala	Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr				
			500					505					510						
Phe	Cys	Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr										
	515					520													
<210> SEQ ID NO 17																			
<211> LENGTH: 535																			
<212> TYPE: PRT																			
<213> ORGANISM: Artificial Sequence																			
<220> FEATURE:																			
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																			
<400> SEQUENCE: 17																			
Gly	Ala	Asp	Asp	Val	Val	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn				
1				5					10					15					
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln				
			20				25						30						
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp				
	35					40						45							
Asp	Trp	Lys	Glu	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly				
	50					55					60								
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val				
65				70					75						80				
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val				
			85					90						95					
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu				
		100						105					110						
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly				
	115					120					125								
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser				
	130				135						140								
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser				
145				150					155					160					
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp				
			165					170					175						
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg				
	180					185						190							
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val				
	195					200					205								
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly				
	210				215						220								
Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu				
225				230					235					240					
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu				
		245						250					255						
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val				
		260					265						270						

-continued

Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val
		275					280					285			
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu
	290					295					300				
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala
305					310					315					320
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser
				325					330					335	
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp
			340					345					350		
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe
	355						360					365			
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His
	370					375					380				
Lys	Thr	Gln	Pro	Phe	Leu	His	Asp	Gly	Tyr	Ala	Val	Ser	Trp	Asn	Thr
385					390					395					400
Val	Glu	Asp	Ser	Ile	Ile	Arg	Thr	Gly	Phe	Gln	Gly	Glu	Ser	Gly	His
				405					410					415	
Asp	Ile	Lys	Ile	Thr	Ala	Glu	Asn	Thr	Pro	Leu	Pro	Ile	Ala	Gly	Val
			420					425					430		
Leu	Leu	Pro	Thr	Ile	Pro	Gly	Lys	Leu	Asp	Val	Asn	Lys	Ser	Lys	Thr
		435					440					445			
His	Ile	Ser	Val	Asn	Gly	Arg	Lys	Ile	Arg	Met	Arg	Cys	Arg	Ala	Ile
	450					455					460				
Asp	Gly	Asp	Val	Thr	Phe	Cys	Arg	Pro	Lys	Ser	Pro	Val	Tyr	Val	Gly
465					470					475					480
Asn	Gly	Val	His	Ala	Asn	Leu	His	Val	Ala	Phe	His	Arg	Ser	Ser	Ser
				485					490					495	
Glu	Lys	Ile	His	Ser	Asn	Glu	Ile	Ser	Ser	Asp	Ser	Ile	Gly	Val	Leu
			500					505					510		
Gly	Tyr	Gln	Lys	Thr	Val	Asp	His	Thr	Lys	Val	Asn	Ser	Lys	Leu	Ser
		515					520					525			
Leu	Phe	Phe	Glu	Ile	Lys	Ser									
	530					535									
<210> SEQ ID NO 18															
<211> LENGTH: 560															
<212> TYPE: PRT															
<213> ORGANISM: Artificial Sequence															
<220> FEATURE:															
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide															
<400> SEQUENCE: 18															
Met	Ser	Arg	Lys	Leu	Phe	Ala	Ser	Ile	Leu	Ile	Gly	Ala	Leu	Leu	Gly
1				5					10				15		
Ile	Gly	Ala	Pro	Pro	Ser	Ala	His	Ala	Gly	Ala	Asp	Asp	Val	Val	Asp
		20						25					30		
Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr
		35					40					45			
Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys
	50					55				60					
Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys	Glu	Phe	Tyr	Ser
65					70					75					80

-continued

Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn	
			85					90					95			
Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val	Thr	Tyr	Pro	Gly	
			100					105					110			
Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys	
		115					120					125				
Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly	
	130					135					140					
Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val	
145					150				155						160	
Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn	
				165				170						175		
Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe	
		180						185					190			
Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala	
	195						200					205				
Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu	
	210					215					220					
Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr	
225					230					235					240	
Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser	
				245				250						255		
Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu	
			260					265					270			
Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu	
	275						280					285				
Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala	
	290					295					300					
Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp	
305					310					315					320	
Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly	
				325					330					335		
Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu	
		340						345					350			
Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala	
		355					360					365				
Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn	
	370					375					380					
Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr	
385					390					395					400	
Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	Gln	Pro	Phe	Leu	His	
				405					410					415		
Asp	Gly	Tyr	Ala	Val	Ser	Trp	Asn	Thr	Val	Glu	Asp	Ser	Ile	Ile	Arg	
			420					425					430			
Thr	Gly	Phe	Gln	Gly	Glu	Ser	Gly	His	Asp	Ile	Lys	Ile	Thr	Ala	Glu	
		435					440					445				
Asn	Thr	Pro	Leu	Pro	Ile	Ala	Gly	Val	Leu	Leu	Pro	Thr	Ile	Pro	Gly	
	450					455					460					
Lys	Leu	Asp	Val	Asn	Lys	Ser	Lys	Thr	His	Ile	Ser	Val	Asn	Gly	Arg	
465					470					475					480	

-continued

Lys	Ile	Arg	Met	Arg	Cys	Arg	Ala	Ile	Asp	Gly	Asp	Val	Thr	Phe	Cys	
			485						490					495		
Arg	Pro	Lys	Ser	Pro	Val	Tyr	Val	Gly	Asn	Gly	Val	His	Ala	Asn	Leu	
			500					505					510			
His	Val	Ala	Phe	His	Arg	Ser	Ser	Ser	Glu	Lys	Ile	His	Ser	Asn	Glu	
		515					520					525				
Ile	Ser	Ser	Asp	Ser	Ile	Gly	Val	Leu	Gly	Tyr	Gln	Lys	Thr	Val	Asp	
	530					535					540					
His	Thr	Lys	Val	Asn	Ser	Lys	Leu	Ser	Leu	Phe	Phe	Glu	Ile	Lys	Ser	
545					550					555					560	
<210> SEQ ID NO 19																
<211> LENGTH: 535																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 19																
Gly	Ala	Asp	Asp	Val	Val	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	
1				5					10					15		
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	
			20					25					30			
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	
		35					40					45				
Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	
	50					55					60					
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	
65					70					75					80	
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	
				85					90					95		
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	
			100					105					110			
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	
		115					120					125				
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	
	130						135				140					
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	
145					150					155					160	
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	
				165					170					175		
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	
		180						185					190			
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	
		195					200					205				
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	
	210					215						220				
Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	
225					230					235					240	
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	
				245					250					255		
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	
		260						265					270			

-continued

Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val
	275						280					285			
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu
	290					295					300				
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala
305					310					315					320
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser
				325					330					335	
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp
			340					345					350		
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe
	355						360					365			
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His
	370					375					380				
Lys	Thr	Gln	Pro	Phe	Phe	His	Asp	Gly	Tyr	Ala	Val	Ser	Trp	Asn	Thr
385					390					395					400
Val	Glu	Asp	Ser	Ile	Ile	Arg	Thr	Gly	Phe	Gln	Gly	Glu	Ser	Gly	His
				405					410					415	
Asp	Ile	Lys	Ile	Thr	Ala	Glu	Asn	Thr	Pro	Leu	Pro	Ile	Ala	Gly	Val
			420					425					430		
Leu	Leu	Pro	Thr	Ile	Pro	Gly	Lys	Leu	Asp	Val	Asn	Lys	Ser	Lys	Thr
		435					440					445			
His	Ile	Ser	Val	Asn	Gly	Arg	Lys	Ile	Arg	Met	Arg	Cys	Arg	Ala	Ile
	450					455					460				
Asp	Gly	Asp	Val	Thr	Phe	Cys	Arg	Pro	Lys	Ser	Pro	Val	Tyr	Val	Gly
465					470					475					480
Asn	Gly	Val	His	Ala	Asn	Leu	His	Val	Ala	Phe	His	Arg	Ser	Ser	Ser
				485					490					495	
Glu	Lys	Ile	His	Ser	Asn	Glu	Ile	Ser	Ser	Asp	Ser	Ile	Gly	Val	Leu
		500						505					510		
Gly	Tyr	Gln	Lys	Thr	Val	Asp	His	Thr	Lys	Val	Asn	Phe	Lys	Leu	Ser
		515					520					525			
Leu	Phe	Phe	Glu	Ile	Lys	Ser									
	530					535									
<210> SEQ ID NO 20															
<211> LENGTH: 560															
<212> TYPE: PRT															
<213> ORGANISM: Artificial Sequence															
<220> FEATURE:															
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide															
<400> SEQUENCE: 20															
Met	Ser	Arg	Lys	Leu	Phe	Ala	Ser	Ile	Leu	Ile	Gly	Ala	Leu	Leu	Gly
1				5					10					15	
Ile	Gly	Ala	Pro	Pro	Ser	Ala	His	Ala	Gly	Ala	Asp	Asp	Val	Val	Asp
		20						25					30		
Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr
		35					40					45			
Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys
	50					55				60					
Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser

-continued

65	70				75				80						
Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn
			85					90					95		
Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val	Thr	Tyr	Pro	Gly
			100					105					110		
Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys
		115					120					125			
Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly
	130					135					140				
Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val
145					150				155						160
Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn
				165				170						175	
Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe
			180					185					190		
Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala
		195					200					205			
Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu
	210					215					220				
Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr
225					230					235					240
Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser
				245					250					255	
Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu
			260					265					270		
Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu
		275					280					285			
Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala
	290					295					300				
Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp
305					310					315					320
Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly
				325					330					335	
Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu
			340					345					350		
Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala
		355					360						365		
Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn
	370					375					380				
Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr
385					390					395					400
Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	Gln	Pro	Phe	Phe	His
				405					410					415	
Asp	Gly	Tyr	Ala	Val	Ser	Trp	Asn	Thr	Val	Glu	Asp	Ser	Ile	Ile	Arg
			420					425					430		
Thr	Gly	Phe	Gln	Gly	Glu	Ser	Gly	His	Asp	Ile	Lys	Ile	Thr	Ala	Glu
		435					440					445			
Asn	Thr	Pro	Leu	Pro	Ile	Ala	Gly	Val	Leu	Leu	Pro	Thr	Ile	Pro	Gly
	450					455					460				
Lys	Leu	Asp	Val	Asn	Lys	Ser	Lys	Thr	His	Ile	Ser	Val	Asn	Gly	Arg
465					470					475					480

His Thr Lys Val Asn Phe Lys Leu Ser Leu Phe Phe Glu Ile Lys Ser
545 550 555 560

<400> SEQUENCE: 21

ggcgagtgct	ttgaatgctt	gggcttcttc	acggcggggc	ttgaccgcgt	tgataagttc	60
gcggccagag	ctgaattgct	ggcgtggggg	ggggttgaac	tggtcgtgtc	ctgccatatc	120
ccttacctgc	tttatcaagt	ctccaaggcg	catcacccgg	ttgtgctgcc	tataccaacg	180
ataagcggta	ggggctttgc	ctgtgtagaa	cgggttgcg	ctaaagcggg	gggaaaagtg	240
cgggtcatgg	tctaaaagct	caccagcac	acgcgtgggt	gctgcaagaa	gcttcactcg	300
cgcagattta	ccgttacggt	cagcgtagac	aggggtcaata	agccatatga	actgggcttt	360
gccgttagtt	gggttaatac	ccaccaggc	tggcccgacg	ctatgagtaa	tcagtgagcg	420
caccacgtcg	cggacgtacg	ggtttaagtc	tgcgggggtca	ccgcctgcgg	tacctacttg	480
gtcaacgtct	acgaccagga	cggcggcgta	ctgcttggtg	gtgagcatgg	cgtactcgca	540
ccgtcctaaa	gcatcagtct	cgaagcgata	catacgcggc	gagttcgtgc	cgtcagcgtt	600
gcgtcgatag	gcctttttta	agtctcgtgt	gactgaaccg	tggagtacat	cgcggcctag	660
atgatcgcgt	aaaaggtcgc	ggtcactggc	agatgctggg	gtgttgtcca	gtccaccacg	720
gtcgcgctcg	acgcgggtag	gtgttttagt	gtgcgcattc	tgcgcagtga	tctgtaaact	780
catgaccgtg	ctttctccca	ggtgtgtgct	gggtgataag	cgaaagtcac	cgggttgccg	840
cccgtgggct	ttcttcgttt	ttcattgtct	ttccctgact	ctaaatgaca	cgggtgttat	900
ttactagcca	tgacacgcga	aaaatatgcc	ttttacctgc	ggttacgtat	ggctagacat	960
atggcaagct	atacgtaacc	gcgtttcagc	tgcacagggc	tgtctgcgca	gatttaccat	1020
cacgggactt	ttcccagttc	aggctgcgca	tatttacgca	tacaacgaaa	gcgggtgcgc	1080
agatttacca	cacactctgc	gctgatttac	cgatacgcag	aaaaagcgtg	cgcagattta	1140
cccatacggg	ggcgaattat	ccagagcaat	aggtatacag	caatacagta	atacagggtc	1200
cataaacctg	tattactgta	ttgctgtatg	cctgtaaacc	tttattttatt	gttgtggacg	1260
tattcttcga	ggtaggtgct	aacaatctcg	cggatgggtca	cgcctttttg	ggcggcgatg	1320
actttaagtt	ctgcgtgaag	gtcgcggctc	atttcaatcg	tcactttctt	gacgtagtcg	1380
cggcctgtgg	gttgggtgaa	tgcgcttcgc	actgttttct	tctcggctgc	tggagttagc	1440
ttcgtgqctt	ttttcattga	ggttcgcqcg	ccttgctqcg	ccctgqcgcg	ttcttttactg	1500

-continued

gtgctcattt catcatctcc atgagttcgt cggcgacgtg gtcgtagccg tgcattgtcgg	1560
ggcctgggca gtatccaaac gctaggtgca tatcttcgcg tagcgggatt tcggttttta	1620
agtgcggcat gtgttccgcg tcgagcgctt ctggtgcgcg gtcaagggcg ctggtgcctt	1680
tcttggcgaa cgtcagtaag actgcatgag gtgttccgtt gactgcgtcg cgcagctccc	1740
atactcggga gaggtcggca gcagcagaac gggtcggaag aatgatgaag tcgctgactg	1800
cgattgctgc ttcgatagcg ttctcgtctc ctggcggcac atcgataccg actgggcat	1860
ggcccactac gtgaaccatc accctaatac agtttttttg ggtcgagggtg ccgtaaagca	1920
ctaaatcgga accctaaagg gagccccga tttagagctt gacggggaaa gccggcgaac	1980
gtggcgagaa aggaaggga gaaagcgaaa ggagcggcg ctagggcgct ggcaagtgt	2040
gcggtcacgc tgcgcgtaac caccacaccc gccgcgctta atgcgcgct acagggcgct	2100
cccatcgcg attcaggctg cgcaactgtt gggaagggcg atcgggtgcg gcctcttcgc	2160
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaacgccag	2220
ggttttccca gtcacgacgt tgtaaacga cggccagtga attgtaatac gactcactat	2280
aggcggaatt ggagctccac cgcgggtggc gccgctctag aactagtga tccagggcat	2340
tgatttcaga gcacccttat aattaggata gctaagtcca ttattttatg agtcctggta	2400
aggggatacg ttgtgagcag aaaactgttt gcgtcaatct taataggggc gctactggg	2460
ataggggccc caccttcagc ccatgcaggc gctgatgatg ttgttgattc ttctaaatct	2520
tttgtgatgg aaaacttttc ttcgtaccac gggactaaac ctggttatgt agattccatt	2580
caaaaaggta tacaaaagcc aaaatctggt acacaaggaa attatgacga tgattggaaa	2640
gggttttata gtaccgacaa taaatacgac gctgcgggat actctgtaga taatgaaaac	2700
ccgctctctg gaaaagctgg aggcgtggc aaagtgcgt atccaggact gacgaagggt	2760
ctcgcactaa aagtggataa tgccgaaact attaagaaag agttaggttt aagtctcact	2820
gaaccgttga tggagcaagt cggaacggaa gaggttatca aaagggttcg tgatgggtgct	2880
tcgctgtag tgctcagcct tcccttcgct gaggggagtt ctagcgttga atatattaat	2940
aactgggaac aggcgaaagc gttaagcgt gaacttgaga ttaattttga aaccgtgga	3000
aaacgtggcc aagatgcgat gtatgagtat atggctcaag cctgtgcagg aaatcgtgtc	3060
aggcgatcag taggtagctc attgtcatgc atcaacctgg attgggatgt tatcgtgat	3120
aaaactaaaa ctaagatcga atctctgaaa gaacacggc cgatcaaaaa caaatgagc	3180
gaaagcccg acaaaactgt atctgaagaa aaagctaaac agtacctgga agaattccac	3240
cagactgcac tggaacaccc ggaactgtct gaacttaaga ccgttactgg taccaaccg	3300
gtattcgtg gtgctaacta cgtgcttg gtagtaaac ttgctcaggt tatcgatagc	3360
gaaactgctg ataactgga aaaaactacc gcggctctgt ctatcctgcc gggtatcgg	3420
agcgtaatgg gcatcgaga cggcgccgtt caccacaaca ctgaagaaat cgttgcacag	3480
tctatcgtc tgagctctct gatgggtgct caggccatcc cgctggtagg tgaactgggt	3540
gatatcgggt tcgctgcata caacttcgtt gaaagcatca tcaacctgtt ccagggtgtt	3600
cacaactctt acaaccgcc ggcttactct ccgggtcaca agacgcagc acctactct	3660
agctctacca agaaaacca gctgcagctc gagcacctgc tgctggattt gcagatgatc	3720
ctgaacggta tcaacaatta caagaacccg aaactgacgc gtatgctgac cttcaagttc	3780

-continued

tacatgccga	agaaggccac	cgaactgaaa	cacctgcagt	gtctagaaga	agaactgaaa	3840
ccgctggagg	aagttctgaa	cctggctcag	tctaaaaact	tccacctgcg	gccgcgtgac	3900
ctgatctcta	acatcaacgt	aatcgttctg	gaactgaagg	gctctgaaac	caccttcatg	3960
tgtgaatacg	ctgatgagac	cgcaaccatc	gtagaattcc	tgaaccgttg	gatcaccttc	4020
tgtcagtcta	tcattctctac	cctgacctga	ggatcccccg	ggctgcagga	attcgatatc	4080
aagcttatcg	ataccgtcga	cctcgagggg	gggcccggta	ccagcttttg	ttccctttag	4140
tgaggggtta	tttcgagctt	ggcgtaatca	tggtcatagc	tgtttctctg	gtgaaattgt	4200
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctggggg	4260
gcctaatacg	tgagtccccg	atccgtcgag	ctcgacctgc	aggggggggg	gggcgctgag	4320
gtctgcctcg	tgaagaaggt	gttgctgact	cataccaggc	ctgaatcgcc	ccatcatcca	4380
gccagaaagt	gagggagcca	cggttgatga	gagctttggt	gtaggtggac	cagttgggtga	4440
ttttgaactt	ttgctttgcc	acggaacggg	ctgcgttgtc	gggaagatgc	gtgatctgat	4500
ccttcaactc	agcaaaaagt	cgatttattc	aacaaagccg	ccgtcccgtc	aagtcagcgt	4560
aatgctctgc	cagtgttaca	accaattaac	caattctgat	tagaaaaact	catcgagcat	4620
caaatgaaac	tgcaatttat	tcatatcagg	attatcaata	ccatattttt	gaaaaagccg	4680
tttctgtaat	gaaggagaaa	actcaccgag	gcagttccat	aggatggcaa	gatectggta	4740
tcggtctgcg	attccgactc	gtccaacatc	aatacaacct	attaatttcc	cctcgtcaaa	4800
aataaggtta	tcaagtgaga	aatcaccatg	agtgacgact	gaatccgggtg	agaatggcaa	4860
aagcttatgc	atttctttcc	agacttggtc	aacaggccag	ccattacgct	cgatcatcaa	4920
atcactcgca	tcaaccaaac	cgttattcat	tcgtgattgc	gcctgagcga	gacgaaatac	4980
gcgatecgtg	ttaaaaggac	aattacaaac	aggaatcgaa	tgcaaccggc	gcaggaacac	5040
tgccagcgca	tcaacaatat	tttcacctga	atcaggatat	tcttctaata	cctggaatgc	5100
tgttttcccg	gggatcgag	tggtgagtaa	ccatgcatca	tcaggagtac	ggataaaatg	5160
cttgatggtc	ggaagaggca	taaattccgt	cagccagttt	agtctgacca	tctcatctgt	5220
aacatcattg	gcaacgctac	ctttgccatg	tttcagaaac	aactctggcg	catcgggctt	5280
cccatacaat	cgatagattg	tcgcacctga	ttgcccagca	ttatcgcgag	cccatttata	5340
cccatataaa	tcagcatcca	tggttgaatt	taatcgcggc	ctcgagcaag	acgtttcccg	5400
ttgaatatgg	ctcataacac	cccttgattt	actgtttatg	taagcagaca	gttttattgt	5460
tcattgatgat	atatttttat	cttggtgcaat	gtaacatcag	agattttgag	acacaacgtg	5520
gctttccccc	ccccccctgc	aggctgagct	cgacggatcg	ggctgcagga	attcggtgag	5580
gttatggcgg	agggttgcga	ggcttaggag	aacagaggaa	gtcatgcttt	gaagcatata	5640
agctgccttg	cccctcaagg	ttttcttcaa	gtgaggtttt	atctaactgc	ctaacggcag	5700
gggaaccgta	tattgcttac	ggtatgagac	cccttaaacg	tccggatagt	caccgctctt	5760
cttttagctcc	gcgacatgcc	tagcaaccgt	ggcgcgagag	actcctacct	ctgcccctat	5820
ttcagcccac	gtgggaactg	tcctgtctcg	gaaatactga	tcgttcacca	tttggctaata	5880
acgggaacttc	gtagatcgtc	cttgagcctt	tttcttaacg	tgcgtctttt	caagcttcga	5940
cctttgtgct	tgcgcataatt	tgcctcggg	gtctgttttc	cagcgttggtg	cggttttttg	6000
tccgcctctg	cgtcccatcg	tggccaaggc	tttccgctcg	ctgctggtgg	ctttacctgg	6060

-continued

tgcgttagag cgcgtgtagg tctcgctctt ggattgggcg acatacccg cgcgcgcct	6120
tgccatgggt tggcggtcgc gcatgggtgg catctcggtg tcgcggcctg caccgccgtg	6180
ggtgtgtgcg acgtttagg cgtgctcata ggcgtcgatg attgctgctg ctgtcaggcg	6240
ttggccttgc tggcgcaagc ggtggccagt cttaagcgca tgtctaaagg ctgtttcgtc	6300
gcgtgctgcg gttccttgga caatccagag cacacgcaca ccgtcgataa gttccgggtc	6360
atactggtcg agaccaccgg cgatttccgc gtctacgtcc tg	6402
<210> SEQ ID NO 22 <211> LENGTH: 25 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide	
<400> SEQUENCE: 22	
Met Ser Arg Lys Leu Phe Ala Ser Ile Leu Ile Gly Ala Leu Leu Gly	
1 5 10 15	
Ile Gly Ala Pro Pro Ser Ala His Ala	
20 25	
<210> SEQ ID NO 23 <211> LENGTH: 6 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic 6xHis tag	
<400> SEQUENCE: 23	
His His His His His His	
1 5	
<210> SEQ ID NO 24 <211> LENGTH: 8 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide	
<400> SEQUENCE: 24	
Asp Tyr Lys Asp Asp Asp Asp Lys	
1 5	
<210> SEQ ID NO 25 <211> LENGTH: 19 <212> TYPE: DNA <213> ORGANISM: Unknown <220> FEATURE: <223> OTHER INFORMATION: Description of Unknown: wild-type toxO sequence	
<400> SEQUENCE: 25	
ttaggatagc tttacctaa	19
<210> SEQ ID NO 26 <211> LENGTH: 32 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence	

<400> SEQUENCE: 26

Asn Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile
20 25 30

<400> SEQUENCE: 27

Ile Gly Ala Pro Pro Ser Ala His Ala Gly Ala Asp Asp Val Val
20 25 30

<400> SEQUENCE: 28

```
<210> SEQ ID NO 29
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide
```

<400> SEQUENCE: 29

```
<210> SEQ ID NO 30
<211> LENGTH: 526
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide
```

<400> SEQUENCE: 30

Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
20 25 30

Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45

-continued

Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	
50						55					60					
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	
65					70					75					80	
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	
				85					90					95		
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	
			100					105					110			
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	
		115					120					125				
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	
	130					135					140					
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	
145					150					155					160	
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	
				165					170					175		
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	
			180					185					190			
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	
		195					200					205				
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	
	210					215						220				
Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	
225					230					235					240	
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	
				245					250					255		
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	
		260						265					270			
Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	
		275					280					285				
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	
	290					295					300					
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	
305					310					315					320	
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	
			325						330					335		
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	
		340						345					350			
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	
	355						360					365				
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	
	370					375					380					
Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	
385					390					395					400	
Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	
			405					410					415			
Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	
		420						425					430			
Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	
	435						440					445				

Gln Ser Ile Ile Ser Thr Leu Thr Glu Asn Leu Tyr Phe Gln
515 520 525

```
<210> SEQ ID NO 31
<211> LENGTH: 1686
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide
```

<400> SEQUENCE: 31

gtgagcagaa	aactgtttgc	gtcaatctta	atagggggcg	tactgggggat	agggggcccca	60
ccttcagccc	atgcaggcgc	tgatgatggt	gctgattcct	ctaaatcttt	tgtgatggaa	120
aacttttctt	cgtaccacgg	gactaaacct	ggttatgtag	attccattca	aaaaggtata	180
caaaagccaa	aatctggtag	acaaggaaat	tatgacgatg	attggaaagg	gttttatagt	240
accgacaata	aatacgacgc	tgcgggatac	tctgtagata	atgaaaaccc	gctctctgga	300
aaagctggag	gcgtgggtcaa	agtgacgtat	ccaggactga	cgaaggttct	cgcactaaaa	360
gtggataatg	cggaaactat	taagaaagag	ttaggtttta	gtctcactga	accgttgatg	420
gagcaagtcg	gaacggaaga	gtttatcaaa	aggttcgggtg	atggtgcttc	gcgtgtagtg	480
ctcagccttc	ccttcgctga	ggggagttct	agcgttgaat	atattaataa	ctgggaacag	540
gcgaaagcgt	taagcgtaga	acttgagatt	aattttgaaa	cccgtggaaa	acgtggccaa	600
gatgcgatgt	atgagtatat	ggctcaagcc	tgtgcaggaa	atcgtgtcag	gcgatcagta	660
ggtagctcat	tgtcatgcat	caacctggat	tgggatgtta	tccgtgataa	aactaaaact	720
aagatcgaat	ctctgaaaga	acacgggtccg	atcaaaaaca	aaatgagcga	aagcccgaac	780
aaaactgtat	ctgaagaaaa	agctaaacag	tacctggaag	aattccacca	gactgcactg	840
gaacaccg	aactgtctga	acttaagacc	gttactggta	ccaaccgggt	attcgtgggt	900
gctaactacg	ctgcttgggc	agtaaacggt	gctcaggtta	tcgatagcga	aactgctgat	960
aacctggaaa	aaactaccgc	ggctctgtct	atcctgccgg	gtatcggtag	cgtaatgggc	1020
atcgcacagc	gcgccgttca	ccacaacact	gaagaaatcg	ttgcacagtc	tatcgctctg	1080
agctctctga	tggttgctca	ggccatcccc	ctggtaggtg	aactggttga	tatcggtttc	1140
gctgcataca	acttcgttga	aagcatcatc	aacctgttcc	aggttgttca	caactcttac	1200
aaccgcccgg	cttactctcc	gggtcacaag	acgcatgcac	ctacttctag	ctctaccaag	1260
aaaaccagc	tgcagctcga	gcacctgctg	ctggatttgc	agatgatcct	gaacggtatc	1320
aacaattaca	agaaccgga	actgacgcgt	atgctgacct	tcaagttcta	catgccgaag	1380
aaggccaccg	aactgaaaca	cctgcagtgt	ctagaagaag	aactgaaacc	gctggaggaa	1440
gttctgaacc	tggtctcagtc	taaaaacttc	cacctgccgc	cgcgtgacct	gatctctaac	1500

-continued

atcaacgtaa tcgttctgga actgaagggc tctgaaacca ctttcatgtg tgaatacgct	1560
gatgagaccg caaccatcgt agaattcctg aaccgttgga tcaccttctg tcagtctatc	1620
atctctaccc tgaccgagaa cctgtacttc cagggccatc accaccacca ccaccatcat	1680
cactag	1686
<210> SEQ ID NO 32 <211> LENGTH: 250 <212> TYPE: PRT <213> ORGANISM: Corynebacterium diphtheriae <400> SEQUENCE: 32	
Met Arg Lys Ile Val Thr Leu Ala Ala Ala Ser Leu Leu Gly Ile Ala	
1 5 10 15	
Gly Thr Ser Gly Val Leu Gly Ala Ala Thr Ala Thr Ala Leu Thr Asn	
20 25 30	
Gly Thr Pro Val Ser Pro Glu Asp Asp Thr Ala Ala Glu Gly Val Val	
35 40 45	
Gln Val Ala Ser Cys Thr Gly Thr Val Val Ala Ser Gln Trp Val Leu	
50 55 60	
Thr Ala Gln His Cys Val Glu Val Pro Asn Leu Gln Arg Pro Val Tyr	
65 70 75 80	
Val Gly Thr Thr Arg Glu Gln Gln Gln Arg Glu Glu Asn Thr Phe Thr	
85 90 95	
Ser Asp Tyr Ala Val Trp Ala Pro His Gly Asp Val Ala Leu Val His	
100 105 110	
Val Thr Asp Ala Leu Pro Gln Arg Leu Val Arg Ala Val Arg Arg Ala	
115 120 125	
Pro Val Ser Phe Gly Glu Gln Gly Arg Val Tyr Gly Trp Gly Ala Gly	
130 135 140	
Thr Gly Glu Thr Leu Gln Tyr Ala Arg Ala Ala Val Gly Lys Thr Ser	
145 150 155 160	
Ser Gly Val Arg Pro Gln Gly Asn Gln His Gly Ala Phe Ile Val Gln	
165 170 175	
Tyr Leu Asp Glu Ala Lys Ala Gly Arg Gly Asp Ser Gly Gly Pro Leu	
180 185 190	
Phe Val Asn Gly Glu Val Ala Gly Val Thr Ser Phe Lys Ala Pro Gln	
195 200 205	
Gly Gly Gly Arg Phe Ser Leu Phe Ala Ser Leu His Gly Leu Gly Asp	
210 215 220	
Trp Ile Ala Gln Thr Thr Ala Ala Lys Pro Glu Asn Pro Asn Ser Lys	
225 230 235 240	
Asn Gln Gln Ser Gln Gln Pro Arg Arg Pro	
245 250	
<210> SEQ ID NO 33 <211> LENGTH: 753 <212> TYPE: DNA <213> ORGANISM: Corynebacterium diphtheriae <400> SEQUENCE: 33	
atgcggaaaa ttgttactct tgctgcagca agcttgcctgg gtattgcggg taccagcggt	60
gttcttggtg cggcaacagc caccgcactg acaaatggca cgcccgctctc ccctgaggac	120

-continued

gataccgctg ccgaaggcgt ggttcaagt gctagttgta ctggcacctg ggttgcttct	180
cagtgggtgc tgaccgcca gcattgcgtc gaggtgcca atttgcagcg gccggtctat	240
gttggcacca cccgcgagca gcaacaacgg gaagaaaata cgttcacctc ggattacgca	300
gtgtgggcac cgcacggcga tgtggcggtg gtgcatgtca ccgatgcgt gccgcagcgt	360
ctgggtccgcg cggttcgccg cgcaccggtg agctttggtg agcagggacg cgtgtacggc	420
tggggtgctg gcaccggcga gacgctgcag tatgcccgcg ctgcggttgg taaaacttct	480
tccggagttc ggccgcaggg caaccagcat ggtgcattca tcgtgcagta tttggacgag	540
gctaaagccg ggcggtggcga ttccggcggg ccgctttttg tcaacggtga ggttgctggg	600
gtgacctcgt ttaaggctcc tcaagggtgg gggcgtttct cgctgtttgc ttcgctgcat	660
gggctaggcg attggattgc gcagaccacc gccgccaagc ccgagaacct gaattccaag	720
aatcaacagt cccagcaacc acgtagaccg tag	753
<210> SEQ ID NO 34	
<211> LENGTH: 1012	
<212> TYPE: DNA	
<213> ORGANISM: Corynebacterium diphtheriae	
<400> SEQUENCE: 34	
ggatccgcaa ttgcgggaat cgctcgtgctt tttagccgaa attttcttgc agccagcgt	60
atgcttatcg cctcagggtt agccgctgac atttaccgct tgaggaataa tatgagtgga	120
aatcaacggt aatttcagaa gacttgccca cacctttagc tagcgaccac cattggtggg	180
cgctagcttt ttgatggctt aagggaacatt tgggcacccg tgtatcgcac attagtcata	240
caggaaatcc tccaagattt cgctccgcatg cccgaccaga cactacagca cccacatagc	300
ttctcgattg tcttgccgag cgggagtagg tagctcacgt gctaccgcac ggggaaccgt	360
atattgctta tgggtgtgcc attaccacc gttggtgcta tgatccgaac ggaaaaagtc	420
agtcgtatta gtgaatcacc gttccgcgcg gcgagaacgc agggctcca caagcgtgtg	480
gtccacaag attgcaagga tgtgtacggt gctggtggcg gtggctccag cctgggtctg	540
tcgtcttctt agcaagtctg cattcacggt tccctaggca atctttgagc aaatccctgt	600
ttaacgcccc tgtacgttcg gcgccgcaga aacctgccgg atcgtgatgt taatcctgcc	660
ttgttccagc ccgcagccgt cgggaagcgt ggcacattc acgcgcacca ccccgatgata	720
agcaaaacgc ttcggccccc cgaaaaaccac caagtcgcgg gagcacagag tcacatcgtc	780
ccagggttgg gtgctgtgatt cgggtgtgtcc catccgaaac agtgcttcgt cgccaatcga	840
tactgaaatg accggcgccc gcgattcttc aaattcatcg acgtgcatgc ccatcccga	900
acccggcgga tagtagttga ccagcaccat ctctgtcacg aaggcctcta cccacggggc	960
tagttcttcg gcaacctccg ctgctgcgcg caacgctgcc ggcgcgggat cc	1012
<210> SEQ ID NO 35	
<211> LENGTH: 242	
<212> TYPE: PRT	
<213> ORGANISM: Corynebacterium diphtheriae	
<400> SEQUENCE: 35	
Met Lys Lys Leu Arg Thr Leu Ala Val Thr Leu Thr Ala Ile Ala Ala	
1 5 10 15	
Ser Thr Met Ala Thr Met Pro Ala Gln Ala Val Ile Ser Pro Thr Pro	

```
<210> SEQ ID NO 36
<211> LENGTH: 729
<212> TYPE: DNA
<213> ORGANISM: Corynebacterium diphtheriae

<400> SEQUENCE: 36

atgaagaaac ttcgtaccct agccgtaacc ctgaccgcaa tgcgagcatt aaccatggcg
accatgccag cacaagcagt gattagtccg acaccgtcac atcaagtttc gctagcgtac
gtcagttttg acaacatgca gtgcaccggc acactcgtca gccccaccgc cgtgctcaca
gcacgacact gcctcaacgg cggcctcggc caggtccgac tcggcgccga tcaacttcacc
gccgtacgtg ccgtggcaca cccccaagca gaccttgccg tcctccacct cgatcgccca
gcgcctaata gcctatccgc aatctctgga cgtcacaccc aaccaggtaa ccgcttcgga
gttgccggct acggaagcac cttcaccggc atccccatgg cagcagctgc aaccatgcaa
cgccgcgtca ccgacgtccc cagccccgac cgccaagcag tcatgatcga aaaccacatc
agccaaggtg tactacgccc aggcgactct ggcgggcccc tcctagaggg caatcacgtc
ataggagtac tcagcatgag cagtgcattc ggccgcgtcg gctggtacat cccaccgca
gaacacgccg actggatcgc ggcggcagcc ggaatccccg caccgggaag cgtcgacaag
cccgtccgc tcgtcgacgc cacagccttc ccgacgcaag agccaagcct cgctagccta
tcctcctag
```

-continued

<210> SEQ ID NO 37
<211> LENGTH: 1012
<212> TYPE: DNA
<213> ORGANISM: Corynebacterium diphtheriae

<400> SEQUENCE: 37

ggatccgggc ttatcaccgc agaagacgcc gaaaaagcca tcgatgccac cctcctagcc 60
gtcgacggca tcacacgcca taacgacgac cccatggcat ggctcgccgc tatgggatac 120
ccactaacat gggcaaaaaa catcacgctt aaggaggccg aatgattacc gtctatcaca 180
acccccgctg ctccacctcg cgcaaagcct tggagtacat agaacaacac agcgacgacg 240
aggtgaccat catccgctac ctcgacgccc ctctagtgga acaagagctg cgcactttgc 300
ttgccgacgc ccacctcagc cgcacgacg caatccgcac caaagaagcc gaatacaaag 360
aactcggact cagctccacc acccccgaat cagagctaata caaggccatg gtcaccacc 420
cacgcctcat tcagcgccca attgtggcaa catgcaaggg aacgcggatc gcccgaccaa 480
ccgaaattct gaaagaaatt ctctagccga atcgccagcg atctcggaag ccaaggaatc 540
cttcggcatg ttaaaaaatg taaaagttta atgcgggcac aaacgcgttg aaagagcaac 600
acaccaccat cggtgaacat acttgattat ttcccacttt ccagaatttt aatgagcatg 660
cccagctcga catcaagaac gcagggaaga tctcactage aatcgacgat aggccttct 720
tcgacaaccc ggaacatctt gccacgtacg acggaactac actggcagga tcgtcaatat 780
gtggtgaatt ccctgttaaa gtaggaagaa cagagggtgt cagtccgttt gtgtacgggg 840
actgacaccc tctgttcaaa tgattcatcg aagacggttt cacttttcta gcaaaacagc 900
aagaccgcga agatttcata tattgggtcaa gcgcagccac aaactatcac tcatagctat 960
gtagccccct attaattcat atctcaaaga gtatccaagc actttgggat cc 1012

<210> SEQ ID NO 38
<211> LENGTH: 558
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 38

Met Ser Arg Lys Leu Phe Ala Ser Ile Leu Ile Gly Ala Leu Leu Gly
1 5 10 15

Ile Gly Ala Pro Pro Ser Ala His Ala Met His His His His His His
20 25 30

Glu Asn Leu Tyr Phe Gln Gly Ala Asp Asp Val Ala Asp Ser Ser Lys
35 40 45

Ser Phe Val Met Glu Asn Phe Ser Ser Tyr His Gly Thr Lys Pro Gly
50 55 60

Tyr Val Asp Ser Ile Gln Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr
65 70 75 80

Gln Gly Asn Tyr Asp Asp Asp Trp Lys Gly Phe Tyr Ser Thr Asp Asn
85 90 95

Lys Tyr Asp Ala Ala Gly Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser
100 105 110

Gly Lys Ala Gly Gly Val Val Lys Val Thr Tyr Pro Gly Leu Thr Lys

-continued

	115					120					125				
Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu
	130					135					140				
Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu
145					150					155				160	
Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu
				165					170					175	
Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu
			180					185					190		
Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg
		195					200					205			
Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys
	210					215					220				
Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile
225					230					235					240
Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu
				245					250					255	
Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro
			260					265					270		
Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe
		275					280					285			
His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val
	290					295					300				
Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala
305					310					315					320
Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu
				325					330					335	
Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met
			340					345					350		
Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala
		355					360					365			
Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu
	370					375					380				
Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu
385					390					395					400
Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro
				405					410					415	
Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr
			420					425					430		
Lys	Lys	Thr	Gln	Leu	Gln	Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met
		435					440					445			
Ile	Leu	Asn	Gly	Ile	Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met
	450					455					460				
Leu	Thr	Phe	Lys	Phe	Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His
465					470					475					480
Leu	Gln	Cys	Leu	Glu	Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn
				485					490					495	
Leu	Ala	Gln	Ser	Lys	Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser
			500					505					510		
Asn	Ile	Asn	Val	Ile	Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe
		515					520					525			

-continued

Met	Cys	Glu	Tyr	Ala	Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn
	530					535					540				
Arg	Trp	Ile	Thr	Phe	Cys	Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr		
545					550					555					
<210> SEQ ID NO 39															
<211> LENGTH: 533															
<212> TYPE: PRT															
<213> ORGANISM: Artificial Sequence															
<220> FEATURE:															
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide															
<400> SEQUENCE: 39															
Met	His	His	His	His	His	His	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Ala	Asp
1				5					10					15	
Asp	Val	Ala	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	Phe	Ser	Ser
			20					25					30		
Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile
		35					40					45			
Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys
	50					55					60				
Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val
65					70					75					80
Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val
				85					90					95	
Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala
			100					105					110		
Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met
		115					120					125			
Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala
	130						135				140				
Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val
145				150						155					160
Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu
			165						170					175	
Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr
		180						185					190		
Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val
	195						200					205			
Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp
	210					215					220				
Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys
225					230					235					240
Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala
				245					250					255	
Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu
		260						265					270		
Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly
		275					280					285			
Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser
	290					295					300				
Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu

-continued																			
305					310					315					320				
Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His				
				325					330					335					
Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met				
			340					345					350						
Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe				
		355					360					365							
Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val				
		370				375					380								
His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	His				
385					390					395					400				
Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	Leu	Glu	His				
			405						410					415					
Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	Asn	Tyr	Lys				
			420					425					430						
Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	Met	Pro	Lys				
		435				440					445								
Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	Glu	Leu	Lys				
		450				455					460								
Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	Phe	His	Leu				
465					470				475						480				
Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	Leu	Glu	Leu				
			485						490					495					
Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	Glu	Thr	Ala				
			500					505					510						
Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	Gln	Ser	Ile				
		515					520					525							
Ile	Ser	Thr	Leu	Thr															
		530																	
<210> SEQ ID NO 40																			
<211> LENGTH: 520																			
<212> TYPE: PRT																			
<213> ORGANISM: Artificial Sequence																			
<220> FEATURE:																			
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																			
<400> SEQUENCE: 40																			
Gly	Ala	Asp	Asp	Val	Ala	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn				
1				5					10					15					
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln				
			20					25					30						
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp				
		35					40					45							
Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly				
	50					55					60								
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val				
65				70						75					80				
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val				
			85					90						95					
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu				
		100						105					110						

Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly		
		115					120					125					
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser		
		130				135					140						
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser		
		145			150					155						160	
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp		
				165					170							175	
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg		
			180					185					190				
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val		
		195					200					205					
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly		
		210				215					220						
Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu		
		225			230					235						240	
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu		
				245					250							255	
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val		
			260					265					270				
Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val		
		275					280					285					
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu		
		290				295					300						
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala		
		305			310					315						320	
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser		
			325					330					335				
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp		
			340					345					350				
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe		
		355					360					365					
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His		
		370				375					380						
Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln		
		385			390					395						400	
Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn		
				405					410							415	
Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr		
			420					425					430				
Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu		
		435					440					445					
Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn		
		450				455					460						
Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val		
		465			470					475						480	
Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp		
				485					490							495	
Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys		
			500					505					510				
Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr										

-continued

515	520	
<210> SEQ ID NO 41 <211> LENGTH: 1677 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 41		
gtgagcagaa aactgtttgc gtcaatctta atagggg'gcgc tactgggggat agggggcccca	60	
ccttcagccc atgcaatgca tcaccaccac caccacgaga acctgtactt ccaggggcgct	120	
gatgatgttg ctgattcttc taaatctttt gtgatggaaa acttttcttc gtaccacggg	180	
actaaacctg gttatgtaga ttccattcaa aaaggtatac aaaagccaaa atctggtaca	240	
caaggaaatt atgacgatga ttggaaaggg ttttatagta ccgacaataa atacgacgct	300	
gcgggatact ctgtagataa tgaaaacccg ctctctggaa aagctggagg cgtggtcaaa	360	
gtgacgtatc caggactgac gaaggttctc gcactaaaag tggataatgc cgaaactatt	420	
aagaaagagt taggtttaag tctcactgaa ccgttgatgg agcaagtcgg aacggaagag	480	
tttatcaaaa gggtcgggtga tgggtgcttcg cgtgtagtgc tcagccttcc cttecgctgag	540	
gggagttcta gcgttgaata tattaataac tgggaacagg cgaaagcgtt aagcgtagaa	600	
cttgagatta attttgaaac ccgtggaaaa cgtggccaag atgcgatgta tgagtatatg	660	
gctcaagcct gtgcaggaaa tcgtgtcagg cgatcagtag gtagctcatt gtcatgcac	720	
aacctggatt gggatgttat ccgtgataaa actaaaacta agatcgaatc tctgaaagaa	780	
cacggtcoga tcaaaaacaa aatgagcgaa agcccgaaca aaactgtatc tgaagaaaaa	840	
gctaaacagt acctggaaga attccaccag actgcactgg aacacccgga actgtctgaa	900	
cttaagaccg ttactggtac caacccggta ttcgctggtg ctaactacgc tgcttgggca	960	
gtaaacgttg ctcaggttat cgatagcgaa actgctgata acctggaaaa aactaccgcg	1020	
gctctgtcta tcctgccggg tatcggtagc gtaatgggca tcgcagacgg cgccggtcac	1080	
cacaacactg aagaaatcgt tgcacagtct atcgctctga gctctctgat gggtgctcag	1140	
gccatcccgc tggtaggtga actggttgat atcggtttcg ctgcatacaa cttegttgaa	1200	
agcatcatca acctgttcca gggtgttcac aactcttaca accgcccggc ttactctccg	1260	
ggtcacaaga cgcatgcacc tacttctagc tctaccaaga aaaccagct gcagctcgag	1320	
cacctgctgc tggatttgca gatgatcctg aacggtatca acaattacaa gaaccgaaa	1380	
ctgacgcgta tgctgacctt caagttctac atgccgaaga aggccaccga actgaaacac	1440	
ctgcagtgtc tagaagaaga actgaaaccg ctggaggaag ttctgaacct ggctcagtct	1500	
aaaaacttcc acctgcggcc gcgtgacctg atctetaaca tcaacgtaat cgttctggaa	1560	
ctgaagggct ctgaaaccac cttcatgtgt gaatacgtg atgagaccgc aaccatcgta	1620	
gaattcctga accgttggat caccttctgt cagtctatca tctctaccct gacctga	1677	
<210> SEQ ID NO 42 <211> LENGTH: 551 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic		

-continued

polypeptide																	
<400> SEQUENCE: 42																	
Met	Ser	Arg	Lys	Leu	Phe	Ala	Ser	Ile	Leu	Ile	Gly	Ala	Leu	Leu	Gly		
1				5					10					15			
Ile	Gly	Ala	Pro	Pro	Ser	Ala	His	Ala	Gly	Ala	Asp	Asp	Val	Ala	Asp		
			20					25					30				
Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	Phe	Ser	Ser	Tyr	His	Gly	Thr		
		35					40					45					
Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys		
	50					55					60						
Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser		
65					70					75					80		
Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn		
			85						90					95			
Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val	Thr	Tyr	Pro	Gly		
			100					105					110				
Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys		
		115					120					125					
Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly		
	130					135					140						
Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val		
145					150					155					160		
Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn		
			165					170						175			
Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe		
		180						185					190				
Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala		
		195					200					205					
Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu		
	210					215					220						
Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr		
225					230					235					240		
Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser		
			245					250						255			
Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu		
		260						265					270				
Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu		
		275					280					285					
Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala		
	290					295					300						
Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp		
305					310					315					320		
Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly		
			325					330						335			
Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu		
		340						345					350				
Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala		
		355					360					365					
Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn		
	370					375					380						

-continued

Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr	
385					390					395					400	
Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	His	Ala	Pro	Thr	Ser	
				405					410					415		
Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	Leu	Glu	His	Leu	Leu	Leu	Asp	
			420					425					430			
Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu	
		435					440					445				
Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	
	450					455					460					
Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	
465					470					475					480	
Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	
				485					490					495		
Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	
			500					505					510			
Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	
		515					520					525				
Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	Gln	Ser	Ile	Ile	Ser	Thr	Leu	
	530					535					540					
Thr	His	His	His	His	His	His										
545					550											
<210> SEQ ID NO 43																
<211> LENGTH: 526																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 43																
Gly	Ala	Asp	Asp	Val	Ala	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	
1				5					10					15		
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	
			20					25					30			
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	
		35					40					45				
Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	
	50					55					60					
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	
65					70					75					80	
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	
			85						90					95		
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	
			100					105					110			
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	
		115					120					125				
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	
						135					140					
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	
145					150					155					160	
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	
				165					170					175		

-continued

Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	
			180					185					190			
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	
		195					200					205				
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	
	210					215					220					
Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	
225					230					235					240	
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	
			245						250					255		
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	
		260						265					270			
Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	
	275						280				285					
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	
	290					295					300					
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	
305					310					315					320	
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	
			325						330					335		
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	
		340						345					350			
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	
	355					360					365					
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	
	370					375					380					
Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	
385					390					395					400	
Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	
		405						410					415			
Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	
		420						425					430			
Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	
	435						440					445				
Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	
	450					455					460					
Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	
465					470					475					480	
Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	
		485						490						495		
Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	
		500						505					510			
Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr	His	His	His	His	His	His			
	515						520					525				

<210> SEQ ID NO 44
<211> LENGTH: 1656
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

-continued

<400> SEQUENCE: 44		
gtgagcagaa aactgtttgc gtcaatctta ataggggcgc tactggggat aggggcccc	60	
ccttcagccc atgcaggcgc tgatgatggt gctgattctt ctaaattctt tgtgatggaa	120	
aacttttctt cgtaccacgg gactaaacct gggtatgtag attccattca aaaagggtata	180	
caaaagccaa aatctggtac acaaggaaat tatgacgatg attggaaagg gttttatagt	240	
accgacaata aatacgacgc tgcgggatac tctgtagata atgaaaaccc gctctctgga	300	
aaagctggag gcggtgtcaa agtgacgtat ccaggactga cgaaggttct cgcactaaaa	360	
gtggataatg ccgaaactat taagaaagag ttaggttta gtctcactga accgttgatg	420	
gagcaagtcg gaacggaaga gtttatcaaa aggttcggtg atgggtgctt cgtgtagt	480	
ctcagccttc ccttcgctga ggggagttct agcgttgaat atattaataa ctgggaacag	540	
gcgaaagcgt taagcgtaga acttgagatt aattttgaaa cccgtggaaa acgtggccaa	600	
gatgcgatgt atgagtatat ggctcaagcc tgtgcaggaa atcgtgtcag gcgatcagta	660	
ggtagctcat tgtcatgcat caacctggat tgggatgtta tccgtgataa aactaaaact	720	
aagatcgaat ctctgaaaga acacggtcct atcaaaaaca aaatgagcga aagcccgaac	780	
aaaactgtat ctgaagaaaa agctaaacag tacctggaag aattccacca gactgcactg	840	
gaacacccgg aactgtctga acttaagacc gttactggta ccaaccgggt attcgctgg	900	
gctaactacg ctgcttgggc agtaaacgtt gctcagggtta tcgatagcga aactgctgat	960	
aacctggaaa aaactaccgc ggctctgtct atcctgccgg gtatcggtag cgtaatgggc	1020	
atcgcagacg ggcgcgttca ccacaacact gaagaaatcg ttgcacagtc tatcgctctg	1080	
agctctctga tggttgctca ggccatcccc ctggttaggtg aactgggtga tatcggttct	1140	
gctgcataca acttcgttga aagcatcatc aacctgttcc aggttggttca caactcttac	1200	
aaccgcccgg cttactctcc gggtcacaag acgcatgcac ctacttctag ctctaccaag	1260	
aaaaccacgc tgcagctcga gcacctgctg ctggatttgc agatgatcct gaacgggtatc	1320	
aacaattaca agaaccgaa actgacgcgt atgtgacct tcaagttcta catgccgaag	1380	
aaggccaccg aactgaaaca cctgcagtgt ctagaagaag aactgaaacc gctggaggaa	1440	
gttctgaacc tggctcagtc taaaaacttc cacctgcggc cgcgtgacct gatctctaac	1500	
atcaacgtaa tcgttctgga actgaagggc tctgaaacca cttcatgtg tgaatacgct	1560	
gatgagaccg caaccatcgt agaattcctg aaccgttggg tcaccttctg tcagtctatc	1620	
atctctaccc tgaccaccca tcaccatcat cactga	1656	
<210> SEQ ID NO 45		
<211> LENGTH: 561		
<212> TYPE: PRT		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide		
<400> SEQUENCE: 45		
Met Ser Arg Lys Leu Phe Ala Ser Ile Leu Ile Gly Ala Leu Leu Gly		
1 5 10 15		
Ile Gly Ala Pro Pro Ser Ala His Ala Gly Ala Asp Asp Val Ala Asp		
20 25 30		
Ser Ser Lys Ser Phe Val Met Glu Asn Phe Ser Ser Tyr His Gly Thr		

-continued																	
35						40						45					
Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	Lys	Gly	Ile	Gln	Lys	Pro	Lys		
50						55					60						
Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	Asp	Trp	Lys	Gly	Phe	Tyr	Ser		
65					70					75					80		
Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	Tyr	Ser	Val	Asp	Asn	Glu	Asn		
				85					90					95			
Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	Val	Lys	Val	Thr	Tyr	Pro	Gly		
			100					105					110				
Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	Asp	Asn	Ala	Glu	Thr	Ile	Lys		
			115				120					125					
Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	Pro	Leu	Met	Glu	Gln	Val	Gly		
	130					135					140						
Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	Asp	Gly	Ala	Ser	Arg	Val	Val		
145					150					155					160		
Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	Ser	Ser	Val	Glu	Tyr	Ile	Asn		
				165					170					175			
Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	Val	Glu	Leu	Glu	Ile	Asn	Phe		
			180					185					190				
Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	Ala	Met	Tyr	Glu	Tyr	Met	Ala		
		195					200					205					
Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	Arg	Ser	Val	Gly	Ser	Ser	Leu		
	210					215					220						
Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	Ile	Arg	Asp	Lys	Thr	Lys	Thr		
225					230					235					240		
Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	Pro	Ile	Lys	Asn	Lys	Met	Ser		
				245					250					255			
Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	Glu	Lys	Ala	Lys	Gln	Tyr	Leu		
			260					265					270				
Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	His	Pro	Glu	Leu	Ser	Glu	Leu		
		275					280					285					
Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	Phe	Ala	Gly	Ala	Asn	Tyr	Ala		
						295					300						
Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	Ile	Asp	Ser	Glu	Thr	Ala	Asp		
305					310					315					320		
Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	Ser	Ile	Leu	Pro	Gly	Ile	Gly		
				325					330					335			
Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	Val	His	His	Asn	Thr	Glu	Glu		
			340					345					350				
Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	Ser	Leu	Met	Val	Ala	Gln	Ala		
			355				360					365					
Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	Ile	Gly	Phe	Ala	Ala	Tyr	Asn		
						375					380						
Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr		
385					390					395					400		
Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	His	Ala	Pro	Thr	Ser		
				405					410					415			
Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	Leu	Glu	His	Leu	Leu	Leu	Asp		
			420					425					430				
Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu		
			435				440					445					

-continued

Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	
450						455					460					
Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	
465					470					475					480	
Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	
				485					490					495		
Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	
			500					505					510			
Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	
		515					520					525				
Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	Gln	Ser	Ile	Ile	Ser	Thr	Leu	
530						535					540					
Thr	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	His	His	His	His	His	His	His	His	
545					550				555						560	
His																
<210> SEQ ID NO 46																
<211> LENGTH: 536																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 46																
Gly	Ala	Asp	Asp	Val	Ala	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn	
1				5					10					15		
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln	
			20					25				30				
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp	
		35				40						45				
Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly	
50						55					60					
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val	
65				70						75					80	
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val	
			85						90					95		
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu	
			100					105					110			
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly	
		115					120					125				
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser	
		130				135					140					
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser	
145					150					155					160	
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp	
			165						170					175		
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg	
			180					185					190			
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val	
		195				200						205				
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly	
		210				215						220				

-continued

Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu	
225					230					235					240	
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu	
				245					250						255	
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val	
			260					265						270		
Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val	
		275					280					285				
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu	
	290					295					300					
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala	
305					310					315					320	
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser	
				325					330						335	
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp	
			340					345						350		
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe	
		355					360						365			
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His	
		370					375					380				
Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln	
385					390					395					400	
Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn	
				405					410						415	
Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr	
			420					425						430		
Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu	
		435					440					445				
Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn	
						455					460					
Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val	
465					470					475					480	
Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp	
				485					490						495	
Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys	
			500					505						510		
Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	His	
		515					520						525			
His	His	His	His	His	His	His	His									
		530				535										
<210> SEQ ID NO 47																
<211> LENGTH: 4679																
<212> TYPE: DNA																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide																
<400> SEQUENCE: 47																
ggatcccccg ggctgcagga attcgatatc aagcttatcg ataccgtcga cctcgagggg															60	
gggcccggtta ccagcttttg ttcccttttag tgagggttaa tttcgagctt ggcgtaatca															120	

-continued

tggtcatagc	tgtttcctgt	gtgaaattgt	tatccgctca	caattccaca	caacatacga	180
gccggaagca	taaagtgtaa	agcctggggt	gcctaattgag	tgagtccccg	atccgtcgag	240
ctcgacctgc	aggggggggg	gggcgctgag	gtctgcctcg	tgaagaaggt	gttgctgact	300
cataccaggc	ctgaatcgcc	ccatcatcca	gccagaaagt	gaggagacca	cggttgatga	360
gagctttggt	gtaggtggac	cagttgggtga	ttttgaactt	ttgctttgcc	acggaacggt	420
ctgcgttgtc	gggaagatgc	gtgatctgat	ccttcaactc	agcaaaagtt	cgattttattc	480
aacaaagccg	ccgtcccgtc	aagtcagcgt	aatgctctgc	cagtgttaca	accaattaac	540
caattctgat	tagaaaaact	catcgagcat	caaatgaaac	tgcaatttat	tcatatcagg	600
attatcaata	ccatattttt	gaaaaagccg	tttctgtaat	gaaggagaaa	actcaccgag	660
gcagttccat	aggatggcaa	gatcctggta	tcggtctgcg	attccgactc	gtccaacatc	720
aatacaacct	attaattttc	cctcgtcaaa	aataagggtta	tcaagtgaga	aatcaccatg	780
agtgacgact	gaatccggtg	agaatggcaa	aagcttatgc	atttctttcc	agacttgttc	840
aacaggccag	ccattacgct	cgtcatcaaa	atcactcgca	tcaaccaaac	cgttattcat	900
tcgtgattgc	gcctgagcga	gacgaaatac	gcgatcgctg	ttaaaaggac	aattacaaac	960
aggaatcgaa	tgcaaccggc	gcaggaacac	tgccagcgca	tcaacaatat	tttcacctga	1020
atcaggatat	tcttctaata	cctggaatgc	tgttttcccg	gggatcgag	tggtgagtaa	1080
ccatgcatca	tcaggagtac	ggataaaatg	cttgatggtc	ggaagaggca	taaattccgt	1140
cagccagttt	agtctgacca	tctcatctgt	aacatcattg	gcaacgctac	ctttgccatg	1200
tttcagaaac	aactctggcg	catcgggctt	cccatacaat	cgatagattg	tcgcacctga	1260
ttgcccagaca	ttatcgcgag	cccatttata	cccatataaa	tcagcatcca	tggtggaatt	1320
taatcgcggc	ctcgagcaag	acgtttcccg	ttgaatatgg	ctcataacac	cccttgattt	1380
actgtttatg	taagcagaca	gttttattgt	tcatgatgat	atatttttat	cttggtgcaat	1440
gtaacatcag	agattttgag	acacaacgtg	gctttccccc	ccccccctgc	aggtegagct	1500
cgacggatcg	ggctgcagga	attcggtgag	gttatggcgg	agggttgcga	ggtctaggag	1560
aacagaggaa	gtcatgcttt	gaagcatata	agctgccctg	cccctcaagg	ttttcttcaa	1620
gtgaggtttt	atctaactgc	ctaacggcag	gggaaccgta	tattgcttac	ggtatgagac	1680
cccttaaacg	tccggatagt	caccgctctt	ctttagctcc	gcgacatgcc	tagcaaccgt	1740
ggcgcgagag	actcctacct	ctgcccctat	ttcagcccac	gtgggaactg	tcctgtctg	1800
gaaatactga	tcgttcacca	tttggttaat	acgggacttc	gtagatcgtc	cttgagcctt	1860
tttcttacgg	tgcgtctttt	caagcttcga	cctttgtgct	tgcgcatatt	tgccctcggg	1920
gtctgttttc	cagcgttggtg	cggttttttg	tccgcctctg	cgtcccatcg	tggtccaaggc	1980
tttccgctcg	ctgctgggtg	ctttacctgg	tgcgtttagag	ccgctgtagg	tctcgctctt	2040
ggattggggc	acatacccg	gcacgcgcct	tgccatgggt	tggtgggtgc	gcattgggtg	2100
catctcggtg	tcgcggcctg	caccgcctg	ggtgtgtgcg	acgttgtagg	cgtgctcata	2160
ggcgctgatg	attgctgcgt	ctgtcaggcg	ttggccttgc	tggtcgcaagc	ggtggccagt	2220
cttaagcgca	tgtctaaagg	ctgtttcgtc	gcgtgctgcg	gttccttgga	caatccagag	2280
cacacgcaca	ccgtcgataa	gttcggggtc	atactggctg	agaccaccgg	cgatttcgcg	2340
gtctacgtcc	tgggcgagtg	ctttgaatgc	ttgggcttct	tcacggcggg	tcttgaccgc	2400

-continued

gttgataagt	tgcgggccag	agctgaattg	ctggcgtggg	gtgggggttg	actggtcgtg	2460
tcttgccata	tcccttacct	gctttatcaa	gtctccaagg	cgcatacccc	ggttgtgctg	2520
cctataccaa	cgataagcgg	taggggcttt	gcctgtgtag	aacgggttgc	ggctaaagcg	2580
gtgggaaaag	tgcgggtcat	ggtctaaaag	ctcaccacgc	acacgcgtgg	ttgctgcaag	2640
aagcttcata	tgcgcagatt	taccgttacg	gtcagcgtag	acaggggtcaa	taagccatat	2700
gaactgggct	ttgccgttag	ttgggttaat	acccacccag	gctggcccga	cgctatgagt	2760
aatcagtgag	cgcaccacgt	cgcggacgta	cgggtttaag	tctgcggggg	caccgcctgc	2820
ggtacctact	tgggtcaacgt	ctacgaccag	gacggcggcg	tactgcttgg	tggtgagcat	2880
ggcgtactcg	caccgtccta	aagcatcagt	ctcgaagcga	tacatacgcg	gcgagttcgt	2940
gccgtcagcg	ttgcgtcgat	aggccttttt	aaagtctcgt	gtgactgaac	cgtggagtac	3000
atcgcgccct	agatgatcgc	gtaaaaggtc	gcggtcactg	gcagatgctg	gggtgttgct	3060
cagtccacca	cggtcgcgct	cgaacggggg	aggtgtttta	gtgtgcgcat	tctgcgcatg	3120
agtctgtaaa	ctcatgaccg	tgtttctctc	cagggtgtgtg	ctgggtgata	agcgaaagtc	3180
atcgggttgc	cgcgcgttgg	ctttcttcgt	ttttcattgt	ctttccctga	ctctaaatga	3240
caccggtggt	atttactagc	catgacacgc	gaaaaatatg	ccttttacct	gcggttacgt	3300
atggctagac	atatggcaag	ctatacgtaa	ccgcgtttca	gctgcacagg	gctgtctgcg	3360
cagatttacc	atcacgggac	ttttcccgat	tcaggctgcg	catatttacg	catacaacga	3420
aagcggttgc	gcagatttac	cacacactct	gcgtgatatt	accgatacgc	agaaaaagcg	3480
tgcgcagatt	taccatacgc	gtggcgaatt	atccagagca	ataggtatac	agcaatacag	3540
taatacaggt	gccataaacc	tgtattactg	tattgctgta	tgcctgtaaa	cctttattta	3600
ttgttggtga	cgtattcttc	gaggtaggtg	ctaacaatct	cgcggatggt	cacgcctttt	3660
tgggcggcga	tgactttaag	ttctgcgtga	aggtcgcggt	cgatttcaat	cgtcatcttc	3720
ttgacgtagt	cgcggcctgt	gggttggtgg	aatgcgcttc	gcactgtttt	cttctcggtc	3780
gctggagtta	gcttcgtggc	ttttttcatt	gaggttcgcg	ggccttgctg	cgccctggcg	3840
cgttctttac	tgggtgctcat	ttcatcatct	ccatgagttc	gtcggcgacg	tggtcgtagc	3900
cgtgcatgtc	ggggcctggg	cagtatccaa	acgctaggtg	catatcttcg	cgtagcggga	3960
tttcggtttt	aaagtgcggc	atgtgttccg	cgtcgagcgc	ttctcgtgcc	gcgtcaaggg	4020
cgtggtgccc	tttctggcgc	aacgtcagta	agactgcattg	aggtgttccg	ttgactgcgt	4080
cgcgcagctc	ccatactcgc	gagaggtcgc	cagcagcaga	acgggtcggg	agaatgatga	4140
agtcgctgac	tgcgattgct	gcttcgatag	cgttctcgtc	tcttgccggc	acatcgatac	4200
cgactgggcg	atggccactc	acgtgaacca	tcaccctaata	caagtttttt	ggggtcgagg	4260
tgcgtaaaag	cactaaatcg	gaaccctaaa	gggagccccc	gatttagagc	ttgacgggga	4320
aagccggcga	acgtggcgag	aaaggaaggg	aagaaagcga	aaggagcggg	cgctagggcg	4380
ctggcaagtg	tagcggtcac	gctgcgcgta	accaccacac	ccgccgcgct	taatgcgccg	4440
ctacagggcg	ctccattcgc	ccattcaggc	tgcgcaactg	ttgggaaggg	cgatcgggtg	4500
gggcctcttc	gctattacgc	cagctggcga	aagggggatg	tgctgcaagg	cgattaagtt	4560
gggtaacgcc	agggttttcc	cagtcacgac	gttgtaaaac	gacggccagt	gaattgtaat	4620
acgactcact	atagggcgaa	ttggagctcc	accgcggtgg	cggccgctct	agaactagt	4679

-continued

<210> SEQ ID NO 48
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
9xHis tag

<400> SEQUENCE: 48

His His His His His His His His
1 5

<210> SEQ ID NO 49
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Description of Unknown:
TEV cleavage site peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: Any small hydrophobic or polar amino acid

<400> SEQUENCE: 49

Glu Asn Leu Tyr Phe Gln Xaa
1 5

<210> SEQ ID NO 50
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 50

His His His His His His Glu Asn Leu Tyr Phe Gln
1 5 10

<210> SEQ ID NO 51
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 51

Gly Ala Asp Asp Val Ala
1 5

<210> SEQ ID NO 52
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 52

Met His His His His His His Glu Asn Leu Tyr Phe Gln
1 5 10

<210> SEQ ID NO 53

<400> SEQUENCE: 53

<210> SEQ ID NO 54

<400> SEQUENCE: 54

<210> SEO ID NO 55

<400> SEQUENCE: 55

<210> SEO ID NO 56

<400> SEQUENCE: 56

<210> SEO ID NO 57

<400> SEQUENCE: 57

<210> SEQ ID NO 58

```
<210> SEQ ID NO 58
<211> LENGTH: 526
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide
```

-continued

<400> SEQUENCE: 58																	
Gly	Ala	Asp	Asp	Val	Val	Asp	Ser	Ser	Lys	Ser	Phe	Val	Met	Glu	Asn		
1				5					10					15			
Phe	Ser	Ser	Tyr	His	Gly	Thr	Lys	Pro	Gly	Tyr	Val	Asp	Ser	Ile	Gln		
			20					25					30				
Lys	Gly	Ile	Gln	Lys	Pro	Lys	Ser	Gly	Thr	Gln	Gly	Asn	Tyr	Asp	Asp		
		35					40					45					
Asp	Trp	Lys	Gly	Phe	Tyr	Ser	Thr	Asp	Asn	Lys	Tyr	Asp	Ala	Ala	Gly		
	50					55					60						
Tyr	Ser	Val	Asp	Asn	Glu	Asn	Pro	Leu	Ser	Gly	Lys	Ala	Gly	Gly	Val		
65					70					75					80		
Val	Lys	Val	Thr	Tyr	Pro	Gly	Leu	Thr	Lys	Val	Leu	Ala	Leu	Lys	Val		
				85					90					95			
Asp	Asn	Ala	Glu	Thr	Ile	Lys	Lys	Glu	Leu	Gly	Leu	Ser	Leu	Thr	Glu		
			100					105					110				
Pro	Leu	Met	Glu	Gln	Val	Gly	Thr	Glu	Glu	Phe	Ile	Lys	Arg	Phe	Gly		
		115					120					125					
Asp	Gly	Ala	Ser	Arg	Val	Val	Leu	Ser	Leu	Pro	Phe	Ala	Glu	Gly	Ser		
	130					135					140						
Ser	Ser	Val	Glu	Tyr	Ile	Asn	Asn	Trp	Glu	Gln	Ala	Lys	Ala	Leu	Ser		
145					150					155				160			
Val	Glu	Leu	Glu	Ile	Asn	Phe	Glu	Thr	Arg	Gly	Lys	Arg	Gly	Gln	Asp		
				165					170					175			
Ala	Met	Tyr	Glu	Tyr	Met	Ala	Gln	Ala	Cys	Ala	Gly	Asn	Arg	Val	Arg		
			180					185					190				
Arg	Ser	Val	Gly	Ser	Ser	Leu	Ser	Cys	Ile	Asn	Leu	Asp	Trp	Asp	Val		
		195				200						205					
Ile	Arg	Asp	Lys	Thr	Lys	Thr	Lys	Ile	Glu	Ser	Leu	Lys	Glu	His	Gly		
	210					215					220						
Pro	Ile	Lys	Asn	Lys	Met	Ser	Glu	Ser	Pro	Asn	Lys	Thr	Val	Ser	Glu		
225					230					235					240		
Glu	Lys	Ala	Lys	Gln	Tyr	Leu	Glu	Glu	Phe	His	Gln	Thr	Ala	Leu	Glu		
			245						250					255			
His	Pro	Glu	Leu	Ser	Glu	Leu	Lys	Thr	Val	Thr	Gly	Thr	Asn	Pro	Val		
			260					265					270				
Phe	Ala	Gly	Ala	Asn	Tyr	Ala	Ala	Trp	Ala	Val	Asn	Val	Ala	Gln	Val		
		275					280					285					
Ile	Asp	Ser	Glu	Thr	Ala	Asp	Asn	Leu	Glu	Lys	Thr	Thr	Ala	Ala	Leu		
	290					295					300						
Ser	Ile	Leu	Pro	Gly	Ile	Gly	Ser	Val	Met	Gly	Ile	Ala	Asp	Gly	Ala		
305					310					315					320		
Val	His	His	Asn	Thr	Glu	Glu	Ile	Val	Ala	Gln	Ser	Ile	Ala	Leu	Ser		
			325						330					335			
Ser	Leu	Met	Val	Ala	Gln	Ala	Ile	Pro	Leu	Val	Gly	Glu	Leu	Val	Asp		
			340					345					350				
Ile	Gly	Phe	Ala	Ala	Tyr	Asn	Phe	Val	Glu	Ser	Ile	Ile	Asn	Leu	Phe		
		355					360					365					
Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg	Pro	Ala	Tyr	Ser	Pro	Gly	His		
	370					375					380						
Lys	Thr	His	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	Gln		

385			390						395					400				
Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	Asn			
			405						410						415			
Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	Tyr			
			420						425						430			
Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	Glu			
			435						440						445			
Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	Asn			
			450						455						460			
Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	Val			
465						470						475			480			
Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	Asp			
			485						490						495			
Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	Cys			
			500						505						510			
Gln	Ser	Ile	Ile	Ser	Thr	Leu	Thr	His	His	His	His	His	His					
515						520						525						

```
<210> SEQ ID NO 59
<211> LENGTH: 1729
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide
```

<400> SEQUENCE: 59

ttgatttcag	agcaccctta	taattaggat	agctaagtcc	attattttat	gagtcctggt	60
aaggggatac	gttgtagaca	gaaaactggt	tgcgtcaatc	ttaatagggg	cgctactggg	120
gataggggcc	ccaccttcag	cccattgcagg	cgtgatgat	gttggtgatt	cttctaaatc	180
ttttgtgatg	gaaaactttt	cttcgtacca	cgggactaaa	cctggttatg	tagattccat	240
tcaaaaaggt	atacaaaagc	caaaatctgg	tacacaagga	aattatgacg	atgattggaa	300
agggttttat	agtaccgaca	ataaatacga	cgtgcggga	tactctgtag	ataatgaaaa	360
cccgtctct	ggaaaagctg	gaggcgtggt	caaagtgcag	tatccaggac	tgacgaaggt	420
tctcgacta	aaagtggata	atgccgaaac	tattaagaaa	gagttagggt	taagtctcac	480
tgaaccgttg	atggagcaag	tcggaacgga	agagtttatc	aaaaggttcg	gtgatggtgc	540
ttcgcgtgta	gtgctcagcc	ttcccttcgc	tgaggggagt	tctagcgttg	aatatattaa	600
taactgggaa	caggcgaaag	cgттаagcgt	agaacttgag	attaattttg	aaaccctggg	660
aaaacgtggc	caagatgcga	tgtatgagta	tatggctcaa	gcctgtgcag	gaaatcgtgt	720
caggcgatca	gtaggtagct	cattgtcatg	catcaacctg	gattgggatg	ttatccgtga	780
taaaactaaa	actaagatcg	aatctctgaa	agaacacggt	ccgatcaaaa	acaaaatgag	840
cgaaagcccg	aacaaaactg	tatctgaaga	aaaagctaaa	cagtacctgg	aagaattcca	900
ccagactgca	ctggaacacc	cggaactgtc	tgaacttaag	accgttactg	gtaccaaccc	960
ggtattcgct	ggtgctaact	acgtcgcttg	ggcagtaaac	gttgctcagg	ttatcgatag	1020
cgaaactgct	gataacctgg	aaaaaactac	cgcggctctg	tctatcctgc	cgggtatcgg	1080
tagcgtaatg	ggcatcgcag	acggcgccgt	tcaccacaac	actgaagaaa	tcgttgacac	1140
gtctatcgct	ctgagctctc	tgatggttgc	tcaggccatc	ccgtcggtag	gtgaactggt	1200

-continued

tgatatcggt ttcgctgcat acaacttcgt tgaaagcatc atcaacctgt tccaggttgt	1260
tcacaactct tacaaccgcc cggtttactc tccgggtcac aagacgcatg cacctacttc	1320
tagctctacc aagaaaaccc agctgcagct cgagcacctg ctgctggatt tgcagatgat	1380
cctgaacggg atcaacaatt acaagaaccc gaaactgacg cgtatgctga ccttcaagtt	1440
ctacatgccg aagaaggcca ccgaactgaa acacctgcag tgtctagaag aagaactgaa	1500
accgctggag gaagttctga acctggctca gtctaaaaac ttccacctgc ggccgcgtga	1560
cctgatctct aacatcaacg taatcgttct ggaactgaag ggctctgaaa ccaccttcac	1620
gtgtgaatac gctgatgaga ccgcaacccat cgtagaattc ctgaaccgtt ggatcacctt	1680
ctgtcagttc atcatctcta cctgaccca ccacacccat catcactga	1729

- What is claimed is:
1. A method for producing aggregate-free, full-length, monomeric diphtheria toxin fusion proteins comprising:

a) transforming bacteria with a DNA expression vector comprising a toxP, a mutant toxO that blocks Fe-mediated regulation of gene expression, and a DNA sequence encoding a protein, wherein the toxP and the mutant toxO regulate expression of the DNA sequence encoding the protein;

b) incubating the transformed bacteria in a culture medium to allow expression of the protein that is secreted into the culture medium; and

c) purifying the protein from the culture medium.

2. The method of claim 1, wherein the bacteria is *Corynebacterium diphtheria*.

3. A method for producing aggregate-free, full-length, monomeric diphtheria toxin fusion proteins comprising:

a) transforming *Corynebacterium diphtheria* strain with a DNA vector comprising:

i) a toxP;

ii) a mutant toxO that blocks Fe-mediated regulation of gene expression;

iii) a DNA sequence encoding a protein comprising:

a) signal peptide;

b) diphtheria receptor binding domain or has a non-functional diphtheria toxin receptor binding domain, and

c) a target receptor binding domain selected from the group comprising IL-2, IL-3, IL-4, IL-6, IL-7, IL-15, EGF, FGF, substance P, CD4, α MSH, GRP,

TT fragment C, GCSF, heregulin β 1, TNF α , TGF β , a functional part thereof, or a combination thereof, wherein the toxP and the mutant toxO regulate expression of the DNA sequence encoding the protein;

b) incubating the transformed bacteria in a culture medium to allow expression of the protein and that is secreted into the culture medium; and

c) purifying the diphtheria toxin fusion protein from the culture medium.

4. The method of claim 3, wherein the diphtheria toxin receptor fusion protein is selected from the any one of SEQ ID NOs:12 to SEQ ID NO:15.

5. The method of claim 3, wherein the *Corynebacterium diphtheria* strain is *Corynebacterium C7* beta (-), tox (-).

6. A method of making a protein comprising:

a) providing a DNA expression vector comprising a toxP, a mutant toxO that blocks Fe-mediated regulation of gene expression, a signal sequence, and a DNA sequence encoding a protein;

b) transforming a bacteria strain with the DNA vector to form a transformant;

c) incubating the transformant in a culture medium for a period of time to allow expression of a protein that is secreted into the culture medium; and

d) purifying the protein from the culture medium.

* * * * *