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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 May 2023 (11.05.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/077210 A1

(51) International Patent Classification:
*C07K 19/00* (2006.01) *C07K 14/195* (2006.01)
*A61K 47/64* (2017.01) *C07K 16/00* (2006.01)
*A61P 35/00* (2006.01) *C07K 16/28* (2006.01)
*C07K 14/35* (2006.01) *C12N 15/62* (2006.01)
*C07K 14/36* (2006.01)

(21) International Application Number:
PCT/CA2022/051225

(22) International Filing Date:
10 August 2022 (10.08.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20210100770 04 November 2021 (04.11.2021) GR

(71) Applicant: **THE HOSPITAL FOR SICK CHILDREN** [CA/CA]; 555 University Avenue, Toronto, Ontario M5G 1X8 (CA).

(72) Inventors: **MELNYK, Roman**; 254 Douglas Avenue, Oakville, Ontario L6J 3S3 (CA). **BEILHARTZ, Greg**; 2335 Vanessa Drive, Burlington, Ontario L7L 7K9 (CA). **GILL, Shivneet**; 38 Banting Crescent, Brampton, Ontario L6Y 2K9 (CA). **SUGIMAN-MARANGOS, Seiji**; 159 Fulton Avenue, Toronto, Ontario M4K 1Y2 (CA).

(74) Agent: **BOOCOCK, Graeme R.** et al.; Borden Ladner Gervais LLP, 100 Queen Street, Suite 1300, Ottawa, Ontario K1P 1J9 (CA).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

(54) Title: NOVEL BACTERIAL TRANSLOCATION DOMAINS AND RECOMBINANT POLYPEPTIDES COMPRISING THEM FOR USE IN CELLULAR DELIVERY

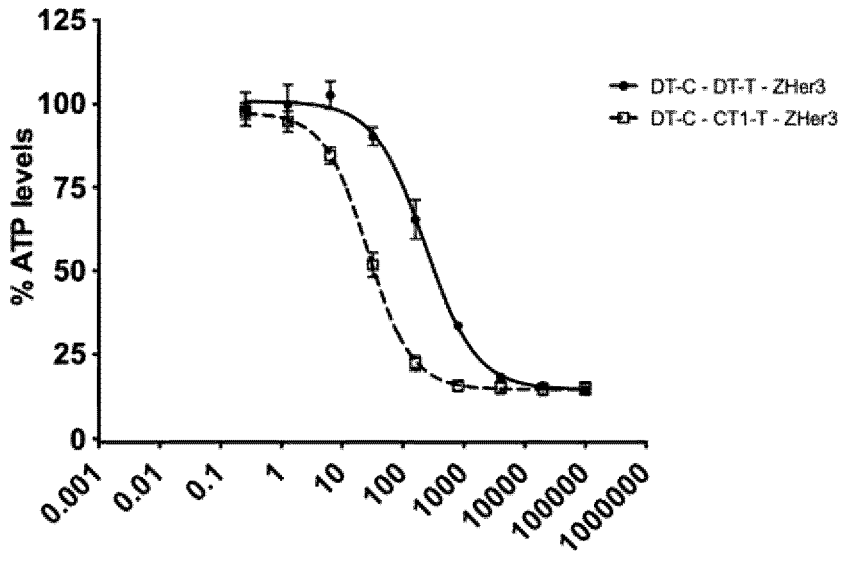



Figure 3E

(57) Abstract: Generally, the present disclosure provides novel bacterial translocation domains for use in cellular delivery. An example is the translocation domain from the Austwickia chelonae protein of SEQ ID NO:2, the translocation domain having the amino acid sequence of SEQ ID NO: 3. Translocase domains from other bacterial strains and species are also described (e.g. SEQ ID NOs: 4 to 17 and 36 to 48). Recombinant polypeptides comprising these translocation domains are described. The recombinant polypeptides are intended for use in delivery of cargo molecules, including therapeutic polypeptides that may be used to treat disease.

WO 2023/077210 A1

*[Continued on next page]*

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

**Declarations under Rule 4.17:**

— *of inventorship (Rule 4.17(iv))*

**Published:**

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

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

# NOVEL BACTERIAL TRANSLOCATION DOMAINS AND RECOMBINANT POLYPEPTIDES COMPRISING THEM FOR USE IN CELLULAR DELIVERY

## CROSS REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims the benefit of priority from Greek Patent Application No. 20210100770 entitled "NOVEL BACTERIAL TRANSLOCATION DOMAINS AND RECOMBINANT POLYPEPTIDES COMPRISING THEM FOR USE IN CELLULAR DELIVERY", which was filed on November 4, 2021.

## FIELD

**[0002]** The present disclosure relates generally to a delivery platform. More particularly, the present disclosure relates to a bacterial toxin-based platform for delivery of cargo molecules to cells.

## BACKGROUND

**[0003]** Therapeutic molecules are often difficult to deliver to cells. They often do not readily penetrate biological membranes. Immunotoxins are a class of biotherapeutics comprised of bacterial toxins, such as diphtheria toxin (DT), that have been repurposed, e.g., into cancer-targeted therapies – both by re-targeting their receptor binding domains (RBD) to target cancer receptors, and by delivering enzyme cargo that target intracellular oncoproteins. However, global vaccination programs against diphtheria has resulted in population-level immunity against DT, and DT-based therapeutics.

**[0004]** There remains a need for delivery platforms that offer the possibility to deliver therapeutics, including protein-based therapeutics, to cells.

## SUMMARY

**[0005]** It is an object of the present disclosure to obviate or mitigate at least one disadvantage of previous approaches.

**[0006]** In a first aspect, the present disclosure provides a recombinant polypeptide of general formula (I):

**[0007]** **A-B-C** (i)

**[0008]** wherein:

**[0009]** **A** is a cargo molecule;

**[0010]** **B** is a translocation polypeptide comprising:

**[0011]** a) a translocation domain from:

**[0012]** the *Austwickia chelonae* protein of SEQ ID NO: 2,

**[0013]** the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1,

**[0014]** the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1,

**[0015]** the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1,

**[0016]** the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1,

**[0017]** the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1,

**[0018]** the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1,

**[0019]** the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1,

**[0020]** the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1,

**[0021]** the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1,

**[0022]** the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1,

**[0023]** the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1,

**[0024]** the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1,

**[0025]** the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1,

**[0026]** the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1;

**[0027]** the *Klebsiella aerogenes* protein of GenBank Accession ElZ2913133.1,

**[0028]** the *Streptomyces sp. MUM 178J* protein of GenBank Accession MCH0551590.1,

**[0029]** the *Crossiella cryophila* protein of GenBank Accession MBB4677777.1,

**[0030]** the *Allokutzneria sp. NRRL B-24872* protein of GenBank Accession WP_143261759.1,

**[0031]** the *Allokutzneria albata* protein of GenBank Accession WP_156051914.1,

**[0032]** the *Streptomyces sp. AV19* protein of GenBank Accession WP_199893204.1,

**[0033]** the *Streptomyces sp. NRBC_110611* protein of GenBank Accession WP_147264604.1,

**[0034]** the *Streptomyces syringium* protein of GenBank Accession WP_209513619.1,

**[0035]** the *Pseudonocardiaceae bacterium YIM PH 21723* protein of GenBank Accession RJQ69589.1,

**[0036]** the *Actinokineospora bangkokensis* protein of GenBank Accession WP_143218892.1,

**[0037]** the *Streptomyces eurocidicus* protein of GenBank Accession MBF6055834.1,

**[0038]** the *Streptomyces pathocidini* protein of GenBank Accession WP_169790908.1, or

**[0039]** the *Streptomyces caatingaensis* protein of GenBank Accession WP_157868472.1, or

**[0040]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a); and

**[0041]** **C** is a targeting moiety.

**[0042]** In one aspect, there is provided a nucleic acid encoding the recombinant polypeptide as defined here.

**[0043]** In one aspect, there is provided a vector comprising the nucleic acid as defined herein.

**[0044]** In further aspect, the present disclosure provides a a composition comprising the recombinant polypeptide as defined herein, together with an acceptable excipient, diluent, or carrier.

**[0045]** In one aspect embodiment, there is provided a pharmaceutical composition the recombinant polypeptide as defined herein, together with a pharmaceutically acceptable excipient, diluent, or carrier.

**[0046]** In one aspect, there is provided a method of delivery a cargo molecule to a cell comprising contacting the cell with the recombinant polypeptide as defined herein.

**[0047]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for delivery of the cargo molecule to a cell.

**[0048]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for preparation of a medicament for delivery of the cargo molecule to a cell.

**[0049]** In one aspect, there is provided the recombinant polypeptide as defined herein for use in delivery of the cargo molecule to a cell.

**[0050]** In one aspect, there is provided a method treating cancer in a subject comprising administering to the subject the recombinant polypeptide as defined herein.

**[0051]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for treatment of cancer in a subject.

**[0052]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for preparation of a medicament for treatment of cancer in a subject.

**[0053]** In one aspect, there is provided the recombinant polypeptide as defined herein for use in treatment of cancer in a subject.

**[0054]** Other aspects and features of the present disclosure will become apparent to those ordinarily skilled in the art upon review of the following description of specific embodiments in conjunction with the accompanying figures.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0055]** Embodiments of the present disclosure will now be described, by way of example only, with reference to the attached Figures.

**[0056]** **Fig. 1A** depicts a schematic of the cell entry mechanism of diphtheria toxin.

**[0057]** **Fig. 1B** depicts the domain organization of DT and CT1 showing the catalytic domain (C) and the bridging furin recognition site (F), followed by the translocase (T) and receptor-binding domains (R).

**[0058]** **Fig. 1C** depicts the crystal structures of diphtheria toxin and CT1.

**[0059]** **Fig. 2A** shows results of experiments assessing a CT1 catalytic domain function and release. DT and a DT-chimera containing the C-domain of CT1 shows similar toxicity on HEK293T cells (b).

**[0060]** **Fig. 2B** shows that DT and a DT-chimera containing the C-domain of CT1 shows no effect on cell viability on $DPH4^{-/-}$ cells.

**[0061]** **Fig. 2C** shows that when DT and CT1 were incubated with cell lysates, there was cleavage at the furin recognition site.

**[0062]** **Fig. 3A** shows that DT and a DT chimeric containing the furin site from CT1 are equipotent on Vero cells, demonstrating efficient cargo release by both toxins.

**[0063]** **Fig. 3B** shows reduced protein synthesis in Vero cells by CT1-T -mediated delivery of DT-C.

**[0064]** **Fig. 3C** shows CT1-T -mediated delivery of the non-native cargo RRSP. RRSP is toxic to RAS mutant cells such as CFPAC-1.

**[0065]** **Fig. 3D** shows that delivery was confirmed by the cleavage of intracellular RAS by RRSP.

**[0066]** **Fig. 3E** shows that CT1-T -mediated delivery of DT-C by measuring cell viability on HPAF II cells. Unexpectedly, CT1-T is a more efficient translocase than DT-T when complexed with targeting domains beyond DT-R.

**[0067]** **Fig. 3F** shows that CT1-T -mediated delivery of RRSP by measuring cell viability on HPAF II cells.

**[0068]** **Fig. 4A** shows results of experiments assessing CT1 human serum binding and neutralization. To quantify the level of pre-existing anti-DT or anti-CT1 antibodies in

human serum, DT or CT1 was immobilized on Nunc MaxiSorp™ plates and incubated with human serum at various dilutions. Wells were then incubated with an anti-human IgG antibody conjugated to HRP, that was developed using TMB reagent. Absorbance was read at 630 nm.

**[0069]** **Fig. 4B** shows results of further experiments assessing CT1 human serum binding and neutralization. To determine the effect of neutralizing antibodies on DT or CT1's capacity to intoxicate cells, various DT/CT1 toxin chimeras (indicated) were incubated with human sera (or PBS) and then added to Vero cells, and protein synthesis levels were measured. EC50 values were calculated and the fold-difference from PBS controls was plotted. Human sera had no effect on the ability of CT1 to intoxicate cells, demonstrating that it is not neutralized by human sera.

**[0070]** **Fig. 5** is a schematic of key residues conserved between DT an CT1.

**[0071]** **Figs. 6A and 6B** depict results of functional characterization of translocases.

**[0072]** **Fig. 7A** shows that antibodies in human sera recognize DT, but show no binding to CT1.

**[0073]** **Fig. 7B** shows that antibodies in human sera recognize the translocase from DT, but not CT1.

**[0074]** **Fig. 8** shows that anti-DT antibodies do not neutralize CT1-based immunotoxins.

**[0075]** **Fig. 9** shows a phylogenetic tree illustrating the relationship of various translocases.

## DETAILED DESCRIPTION

**[0076]** Generally, the present disclosure provides novel bacterial translocation domains for use in cellular delivery. Recombinant polypeptides comprising these translocation domains are described. The recombinant polypeptides are intended for use in delivery of cargo molecules, including therapeutic polypeptides.

**[0077]** **Recombinant Polypeptides**

**[0078]** In one aspect, there is provided a recombinant polypeptide of general formula (I):

**[0079]** **A-B-C** (i)

**[0080]** wherein:

**[0081]** **A** is a cargo molecule;

**[0082]** **B** is a translocation polypeptide comprising:

**[0083]** a) a translocation domain from:

**[0084]** the *Austwickia chelonae* protein of SEQ ID NO: 2,

**[0085]** the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1,

**[0086]** the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1,

**[0087]** the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1,

**[0088]** the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1,

**[0089]** the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1,

**[0090]** the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1,

**[0091]** the *Streptomyces MBT76* protein of GenBank Accession WP_0791103f21.1,

**[0092]** the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1,

**[0093]** the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1,

**[0094]** the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1,

**[0095]** the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1,

**[0096]** the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1,

**[0097]** the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1,

**[0098]** the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1;

**[0099]** the *Klebsiella aerogenes* protein of GenBank Accession EIZ2913133.1,

**[00100]** the *Streptomyces sp. MUM 178J* protein of GenBank Accession MCH0551590.1,

**[00101]** the *Crossiella cryophila* protein of GenBank Accession MBB4677777.1,

**[00102]** the *Allokutzneria sp. NRRL B-24872* protein of GenBank Accession WP_143261759.1,

**[00103]** the *Allokutzneria albata* protein of GenBank Accession WP_156051914.1,

**[00104]** the *Streptomyces sp. AV19* protein of GenBank Accession WP_199893204.1,

**[00105]** the *Streptomyces sp. NRBC_110611* protein of GenBank Accession WP_147264604.1,

**[00106]** the *Streptomyces syringium* protein of GenBank Accession WP_209513619.1,

**[00107]** the *Pseudonocardiaceae bacterium YIM PH 21723* protein of GenBank Accession RJQ69589.1,

**[00108]** the *Actinokineospora bangkokensis* protein of GenBank Accession WP_143218892.1,

**[00109]** the *Streptomyces eurocidicus* protein of GenBank Accession MBF6055834.1,

**[00110]** the *Streptomyces pathocidini* protein of GenBank Accession WP_169790908.1, or

**[00111]** the *Streptomyces caatingaensis* protein of GenBank Accession WP_157868472.1, or

**[00112]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a); and

**[00113]** **C** is a targeting moiety.

**[00114]** In one embodiment, a), the translocation domain is from:

**[00115]** - the *Austwickia chelonae* protein of SEQ ID NO: 2,

**[00116]** - the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1,

**[00117]** - the *Streptomyces sp.*TLI 053 protein of GenBank Accession SDT83331.1,

**[00118]** - the *Streptomyces sp.* SLBN-118 protein of GenBank Accession WP_160159328.1,

**[00119]** - the *Streptomyces sp.* AA8 protein of GenBank Accession WP_168096531.1,

**[00120]** - the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1,

**[00121]** - the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1,

**[00122]** - the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1,

**[00123]** - the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1,

**[00124]** - the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1,

**[00125]** - the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1,

**[00126]** - the *Austwickia sp.* TVS 96-490-7B protein of GenBank Accession WP_219106995.1, or

**[00127]** - the *Austwickia chelonae* LK16-18 protein of GenBank Accession WP_162873017.1.

**[00128]** In one embodiment, in a), the translocation domain is from:

**[00129]** - the *Austwickia chelonae* protein of SEQ ID NO: 2,

**[00130]** - the *Streptomyces sp.*TLI 053 protein of GenBank Accession SDT83331.1,

**[00131]** - the *Streptomyces* klenkii protein of GenBank Accession WP_120757473.1,

**[00132]** - the *Austwickia sp.* TVS 96-490-7B protein of GenBank Accession WP_219106995.1, or

**[00133]** - the *Austwickia chelonae* LK16-18 protein of GenBank Accession WP_162873017.1.

**[00134]** In one embodiment, in a), the translocation domain is from:

**[00135]** - the *Austwickia chelonae* protein of SEQ ID NO: 2,

**[00136]** - the *Austwickia chelonae* LK16-18 protein of GenBank Accession WP_162873017.1.

**[00137]** In one embodiment, in a):

**[00138]** i. the translocation domain from the *Austwickia chelonae* protein herein named CT1 has the amino acid sequence of SEQ ID NO: 3,

**[00139]** ii. the translocation domain from the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1 has the amino acid sequence of SEQ ID NO: 4,

**[00140]** iii. the translocation domain from the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1 has the amino acid sequence of SEQ ID NO: 5,

**[00141]** iv. the translocation domain from the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1 has the amino acid sequence of SEQ ID NO: 6,

**[00142]** v. the translocation domain from the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1 has the amino acid sequence of SEQ ID NO: 7,

**[00143]** vi. the translocation domain from the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1 has the amino acid sequence of SEQ ID NO: 8,

**[00144]** vii. the translocation domain from the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1 has the amino acid sequence of SEQ ID NO: 9,

**[00145]** viii. the translocation domain from the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1 has the amino acid sequence of SEQ ID NO: 10,

**[00146]** ix. the translocation domain from the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1 has the amino acid sequence of SEQ ID NO: 11,

**[00147]** x. the translocation domain from the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1 has the amino acid sequence of SEQ ID NO: 12,

**[00148]** xi. the translocation domain from the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1 has the amino acid sequence of SEQ ID NO: 13,

**[00149]** xii. the translocation domain from the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1 has the amino acid sequence of SEQ ID NO: 14,

**[00150]** xiii. the translocation domain from the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1 has the amino acid sequence of SEQ ID NO: 15,

**[00151]** xiv. the translocation domain from the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1 has the amino acid sequence of SEQ ID NO: 16,

**[00152]** xv. the translocation domain from the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1 has the amino acid sequence of SEQ ID NO: 17, and

**[00153]** xvi. the translocation domain has the amino acid sequence of any one of SEQ ID Nos: 36 to 48.

**[00154]** In one embodiment, in a):

**[00155]** i. the translocation domain from the Austwickia chelonae protein herein named CT1 has the amino acid sequence of SEQ ID NO: 3,

**[00156]** ii. the translocation domain from the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1 has the amino acid sequence of SEQ ID NO: 4,

**[00157]** iii. the translocation domain from the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1 has the amino acid sequence of SEQ ID NO: 5,

**[00158]** iv. the translocation domain from the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1 has the amino acid sequence of SEQ ID NO: 6,

**[00159]** v. the translocation domain from the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1 has the amino acid sequence of SEQ ID NO: 7,

**[00160]** vi. the translocation domain from the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1 has the amino acid sequence of SEQ ID NO: 8,

**[00161]** vii. the translocation domain from the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1 has the amino acid sequence of SEQ ID NO: 9,

**[00162]** viii. the translocation domain from the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1 has the amino acid sequence of SEQ ID NO: 10,

**[00163]** ix. the translocation domain from the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1 has the amino acid sequence of SEQ ID NO: 11,

**[00164]** xi. the translocation domain from the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1 has the amino acid sequence of SEQ ID NO: 13,

**[00165]** xiii. the translocation domain from the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1 has the amino acid sequence of SEQ ID NO: 15,

**[00166]** xiv. the translocation domain from the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1 has the amino acid sequence of SEQ ID NO: 16, and

**[00167]** xv. the translocation domain from the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1has the amino acid sequence of SEQ ID NO: 17.

**[00168]** A "**translocation polypeptide**", as referred to herein, is intended to refer to a polypeptide that comprises a translocation domain.

**[00169]** A "**translocation domain**" as referred to herein is a polypeptide sequence that functions to facilitate transport the protein in which it occurs across a cell membrane, thereby facilitating cell entry. This activity can be assessed, for example, using the assays described herein.

**[00170]** In one embodiment, the translocation polypeptide comprises:

**[00171]** a) the translocation domain from the *Austwickia chelonae* protein, CT1, wherein the translocation domain has amino acid sequence of SEQ ID NO: 3 (herein termed "**CT1-T**"), or

**[00172]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00173]** In one embodiment, the translocation polypeptide comprises:

**[00174]** a) the translocation domain from the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 4, or

**[00175]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00176]** In one embodiment, the translocation polypeptide comprises:

**[00177]** a) the translocation domain from the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 5, or

**[00178]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00179]** In one embodiment, the translocation polypeptide comprises:

**[00180]** a) the translocation domain from the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 6, or

**[00181]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00182]** In one embodiment, the translocation polypeptide comprises:

**[00183]** a) the translocation domain from the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 7, or

**[00184]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00185]** In one embodiment, the translocation polypeptide comprises:

**[00186]** a) the translocation domain from the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 8, or

**[00187]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00188]** In one embodiment, the translocation polypeptide comprises:

**[00189]** a) the translocation domain from the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 9, or

**[00190]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00191]** In one embodiment, the translocation polypeptide comprises:

**[00192]** a) the translocation domain from the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 10, or

**[00193]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00194]** In one embodiment, the translocation polypeptide comprises:

**[00195]** a) the translocation domain from the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 11, or

**[00196]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00197]** In one embodiment, the translocation polypeptide comprises:

**[00198]** a) the translocation domain from the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 12, or

**[00199]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00200]** In one embodiment, the translocation polypeptide comprises:

**[00201]** a) the translocation domain from the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 13, or

**[00202]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00203]** In one embodiment, the translocation polypeptide comprises:

**[00204]** a) the translocation domain from the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 14, or

**[00205]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00206]** the translocation polypeptide comprises:

**[00207]** a) the translocation domain from the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 15, or

**[00208]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00209]** In one embodiment, the translocation polypeptide comprises:

**[00210]** a) the translocation domain from the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 16, or

**[00211]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00212]** In one embodiment, the translocation polypeptide comprises:

**[00213]** a) the translocation domain from the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 17, or

**[00214]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00215]** In one embodiment, the translocation polypeptide comprises:

**[00216]** a) the translocation domain from the *Klebsiella aerogenes* protein of GenBank Accession EIZ2913133.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 36, or

**[00217]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00218]** In one embodiment, the translocation polypeptide comprises:

**[00219]** a) the translocation domain from the *Streptomyces sp. MUM 178J* protein of GenBank Accession MCH0551590.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 37, or

**[00220]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00221]** In one embodiment, the translocation polypeptide comprises:

**[00222]** a) the translocation domain from the *Crossiella cryophila* protein of GenBank Accession MBB4677777.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 38, or

**[00223]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00224]** In one embodiment, the translocation polypeptide comprises:

**[00225]** a) the translocation domain from the *Allokutzneria sp. NRRL B-24872* protein of GenBank Accession WP_143261759.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 39, or

**[00226]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00227]** In one embodiment, the translocation polypeptide comprises:

**[00228]** a) the translocation domain from the *Allokutzneria albata* protein of GenBank Accession WP_156051914.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 40, or

**[00229]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00230]** In one embodiment, the translocation polypeptide comprises:

**[00231]** a) the translocation domain from the *Streptomyces sp. AV19* protein of GenBank Accession WP_199893204.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 41, or

**[00232]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00233]** In one embodiment, the translocation polypeptide comprises:

**[00234]** a) the translocation domain from the *Streptomyces sp. NRBC_110611* protein of GenBank Accession WP_147264604.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 42, or

**[00235]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00236]** In one embodiment, the translocation polypeptide comprises:

**[00237]** a) the translocation domain from the *Streptomyces syringium* protein of GenBank Accession WP_209513619.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 43, or

**[00238]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00239]** In one embodiment, the translocation polypeptide comprises:

**[00240]** a) the translocation domain from the *Pseudonocardiaceae bacterium YIM PH 21723* protein of GenBank Accession RJQ69589.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 44, or

**[00241]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00242]** In one embodiment, the translocation polypeptide comprises:

**[00243]** a) the translocation domain from the *Actinokineospora bangkokensis* protein of GenBank Accession WP_143218892.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 45, or

**[00244]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00245]** In one embodiment, the translocation polypeptide comprises:

**[00246]** a) the translocation domain from the *Streptomyces eurocidicus* protein of GenBank Accession MBF6055834.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 46, or

**[00247]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00248]** In one embodiment, the translocation polypeptide comprises:

**[00249]** a) the translocation domain from the *Streptomyces pathocidini* protein of GenBank Accession WP_169790908.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 47, or

**[00250]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00251]** In one embodiment, the translocation polypeptide comprises:

**[00252]** a) the translocation domain from the *Streptomyces caatingaensis* protein of GenBank Accession WP_157868472.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 48, or

**[00253]** b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

**[00254]** The *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1 may be referred to herein as "CT2". The *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1 may be referred to herein as "CT3".

**[00255]** In one embodiment, the translocation domain is as defined in b) in any one of the above embodiments and is at least at least 85% identical to the translocation domain defined in a) across the full length thereof. In one embodiment, the translocation domain is as defined in b) in any one of the above embodiments and is at least at least 90% identical to the translocation domain defined in a) across the full length thereof. In one embodiment, the translocation domain is as defined in b) in any one of the above embodiments and is at least at least 95% identical to the translocation domain defined in a) across the full length thereof. In one embodiment, the translocation domain is as defined in b) in any one of the above embodiments and is at least at least 98% identical to the translocation domain defined in a) across the full length thereof. In one embodiment, the translocation domain is as defined in b) in any one of the above embodiments and is at least at least 99% identical to the translocation domain defined in a) across the full length thereof.

**[00256]** In one embodiment, the translocation polypeptide is as defined in a) in any one of the above embodiments.

**[00257]** In one embodiment, the translocation polypeptide comprises:

**[00258]** a) the translocation domain from the *Austwickia chelonae* CT1 protein, wherein the translocation domain has amino acid sequence of SEQ ID NO: 3.

**[00259]** In one embodiment, the translocation polypeptide comprises:

**[00260]** a) the translocation domain from the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 4.

**[00261]** In one embodiment, the translocation polypeptide comprises:

**[00262]** a) the translocation domain from the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 5.

**[00263]** In one embodiment, the translocation polypeptide comprises:

**[00264]** a) the translocation domain from the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 6.

**[00265]** In one embodiment, the translocation polypeptide comprises:

**[00266]** a) the translocation domain from the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 7.

**[00267]** In one embodiment, the translocation polypeptide comprises:

**[00268]** a) the translocation domain from the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1.

**[00269]** In one embodiment, the translocation polypeptide comprises:

**[00270]** a) the translocation domain from the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1.

**[00271]** In one embodiment, the translocation polypeptide comprises:

**[00272]** a) the translocation domain from the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 10.

**[00273]** In one embodiment, the translocation polypeptide comprises:

**[00274]** a) the translocation domain from the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1.

**[00275]** In one embodiment, the translocation polypeptide comprises:

**[00276]** a) the translocation domain from the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 12.

**[00277]** In one embodiment, the translocation polypeptide comprises:

**[00278]** a) the translocation domain from the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 13.

**[00279]** In one embodiment, the translocation polypeptide comprises:

**[00280]** a) the translocation domain from the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 14.

**[00281]** the translocation polypeptide comprises:

**[00282]** a) the translocation domain from the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 15.

**[00283]** In one embodiment, the translocation polypeptide comprises:

**[00284]** a) the translocation domain from the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 16.

**[00285]** In one embodiment, the translocation polypeptide comprises:

**[00286]** a) the translocation domain from the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 17.

**[00287]** In some embodiments, the translocation domains having the amino acid sequences of SEQ ID NO: 12 (from the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1) and 14 (from the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1) are excluded from the above-described embodiments.

**[00288]** In some embodiments, the translocation polypeptide may comprise functional truncations of the full-length protein that comprise any one of the above-describe corresponding translocation domains, wherein the function of the translocation domain is maintained.

**[00289]** In one embodiment, A and B are separated by a linker. In one embodiment, A and B are separated by an amino acid linker. In one embodiment, the amino acid linker comprises $(G_4S)_2$. In one embodiment, the linker is cleavable. In one embodiment, the linker comprises a protease recognition site. In one embodiment, the protease recognition site isa furin protease recognition site. In one embodiment, the protease recognition site is bracketed by cysteine residues to allow for disulphide bond formation to form an intramolecular loop. In one embodiment, the amino acid linker comprises a furin protease recognition site bracketed by cysteine residues. In one example embodiment, the amino acid linker comprises SEQ ID NO: 32, which comprises the furin protease recognition site and bracketing cysteine residues. In one embodiment, said $(G_4S)_2$ is positioned N-terminally with respect to said SEQ ID NO: 32. In one embodiment, the linker is self-cleaving. In one embodiment, the linker is self-clearing. In one embodiment, the linker comprises an autoprocessing domain.

**[00290]** In one embodiment, B and C are separated by a linker. In one embodiment, B and C are separated by an amino acid linker. In one embodiment, the amino acid linker comprises (G4S)2. In one embodiment, the amino acid linker comprises (G4S)2. In one embodiment, the amino acid linker comprises SEQ ID NO: 33. In one embodiment, wherein said (G4S)2 is positioned N-terminally with respect to said SEQ ID NO: 33.

**[00291]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 821 of SEQ ID NO: 22, preferably 90% identical to amino acids 1 to 821 of SEQ ID NO: 22, more preferably 95% identical to amino acids 1 to 821 of SEQ ID NO: 22, even more preferably 100% identical to amino acids 1 to 821 of SEQ ID NO: 22.

**[00292]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 22.

**[00293]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 822 of SEQ ID NO: 23, preferably 90% identical to amino acids 1 to 822 of SEQ ID NO: 23, more preferably 95% identical to amino acids 1 to 822 of SEQ ID NO: 23, even more preferably 100% identical to amino acids 1 to 822 of SEQ ID NO: 23.

**[00294]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 23.

**[00295]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 822 SEQ ID NO: 24, preferably 90% identical to amino acids 1 to 822 SEQ ID NO: 24, more preferably 95% identical to amino acids 1 to 822 SEQ ID NO: 24, even more preferably 100% identical to amino acids 1 to 822 SEQ ID NO: 24.

**[00296]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 24.

**[00297]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 821 of SEQ ID NO: 25, preferably 90% identical to amino acids 1 to 821 of SEQ ID NO: 25, more preferably 95% identical to amino acids 1 to 821 of SEQ ID NO: 25, even more preferably 100% identical to amino acids 1 to 821 of SEQ ID NO: 25.

**[00298]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 25.

**[00299]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 806 of SEQ ID NO: 26, preferably 90% identical to amino acids 1 to 806 of SEQ ID NO: 26, more preferably 95% identical to amino acids 1 to 806 of SEQ ID NO: 26, even more preferably 100% identical to amino acids 1 to 806 of SEQ ID NO: 26.

**[00300]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 26.

**[00301]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 811 of SEQ ID NO: 27, preferably 90% identical to amino acids 1 to 811 of SEQ ID NO: 27, more preferably 95% identical to amino acids 1 to 811 of SEQ ID NO: 27, even more preferably 100% identical to amino acids 1 to 811 of SEQ ID NO: 27.

**[00302]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 27.

**[00303]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 822 of SEQ ID NO: 28, preferably 90% identical to amino acids 1 to 822 of SEQ ID NO: 28, more preferably 95% identical to amino acids 1 to 822 of SEQ ID NO: 28, even more preferably 100% identical to amino acids 1 to 822 of SEQ ID NO: 28.

**[00304]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 28.

**[00305]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 796 of SEQ ID NO: 29, preferably 90% identical to amino acids 1 to 796 of SEQ ID NO: 29, more preferably 95% identical to amino acids 1 to 796 of SEQ ID NO: 29, even more preferably 100% identical to amino acids 1 to 796 of SEQ ID NO: 29.

**[00306]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 29.

**[00307]** In one embodiment, the recombinant polypeptide comprises an amino acid sequence that is at least 80% identical to amino acids 1 to 822 of SEQ ID NO: 30, preferably 90% identical to amino acids 1 to 822 of SEQ ID NO: 30, more preferably 95% identical to amino acids 1 to 822 of SEQ ID NO: 30, even more preferably 100% identical to amino acids 1 to 822 of SEQ ID NO: 30.

**[00308]** In one embodiment, the recombinant polypeptide comprises the sequence of SEQ ID NO: 30.

**[00309]** In one embodiment, the targeting moiety comprises a targeting polypeptide or aptamer.

**[00310]** In one embodiment, the targeting polypeptide comprises an antibody, a binding fragment of an antibody, an affibody, an affitin, a DARPin, or a receptor ligand.

**[00311]** In one embodiment, the targeting polypeptide comprises an affibody against Her3. In one embodiment, the affibody against Her3 comprises the amino acid sequence of SEQ ID NO: 19.

**[00312]** In one embodiment, targeting polypeptide comprises a receptor ligand for $\alpha v\beta 6$ integrin. In one embodiment, the receptor ligand for $\alpha v\beta 6$ integrin comprises the amino acid sequence of SEQ ID NO: 20.

**[00313]** In one embodiment, the targeting moiety comprises at least two targeting polypeptides, at least two aptamers, or a combination of a targeting polypeptide and an aptamer. In one embodiment, the at least two targeting polypeptides are selected from the group an antibody, a binding fragment of an antibody, an affibody, a peptide, an affitin, a DARPin, a receptor ligand, and combinations thereof. In one embodiment, the at least two

targeting polypeptides comprise an affibody against Her3. In one embodiment, the affibody against Her3 comprises the amino acid sequence of SEQ ID NO: 19. In one embodiment, the at least two targeting polypeptides comprise a receptor ligand for αvβ6 integrin. In one embodiment, the receptor ligand for αvβ6 integrin comprises the amino acid sequence of SEQ ID NO: 20. In one embodiment, the at least two targeting polypeptides comprise both the affibody against Her3 and the receptor ligand for αvβ6 integrin, preferably wherein the former comprises the amino acid sequence of SEQ ID NO: 19 and the latter comprises the amino acid sequence of SEQ ID NO: 20. In one embodiment, the at least two least two targeting polypeptides, the at least two aptamers, or the combination are separated by an amino acid linker. In one embodiment, the amino acid linker comprises $(G_4S)_2$.

**[00314]** In one embodiment, the targeting moiety binds to a cell surface protein.

**[00315]** In one embodiment, the cell surface protein is lineage-specific or tissue-specific.

**[00316]** In one embodiment, the cell surface protein is ubiquitously expressed.

**[00317]** In one embodiment, the cell surface protein is expressed in a disease cell.

**[00318]** In one embodiment, the cell surface protein is specific to a disease cell and is not expressed in a corresponding healthy cell.

**[00319]** In one embodiment, the cell surface protein has elevated expression in a disease cell compared to a corresponding healthy cell.

**[00320]** In one embodiment, the disease cell is a cancer cell.

**[00321]** In one embodiment, the cargo molecule comprises a cargo polypeptide.

**[00322]** The cargo polypeptide may comprise any polypeptide for which cellular delivery is desired. The cargo polypeptide may comprise an enzyme, or an active fragment thereof having substantially the same activity. By 'substantially the same activity' is meant that a core function of the enzyme is substantially unaltered in the fragment.

**[00323]** The cargo polypeptide may have a molecular weight of less than 10 kDa, greater than 10 kDa, greater than 20 kDa, greater than 30 kDa, greater than 50 kDa, greater than 100 kDa, or greater than 150 kDa.

**[00324]** The cargo polypeptide comprises a genome-modifying protein. The genome-modifying protein comprises a zinc finger nuclease (ZFN), a transcription activator-like effector nuclease (TALEN), or a CRISPR (clustered regularly interspaced short palindromic repeat) protein. The CRISPR protein may be Cas9. The cargo polypeptide may comprise a complex of the genome-modifying protein and a nucleic acid, such as a guide nucleic acid. For instance, Cas9 may be complexed with a nucleic acid (such as a guide RNA), such as crRNA, trRNA, and/or sgRNA.

**[00325]** In one embodiment, the cargo molecule comprises a therapeutic polypeptide.

**[00326]** By '**therapeutic polypeptide'** is meant any protein, the cellular delivery of which could be used for a therapeutic purpose. It is well known, for example, that many human diseases or disorders are caused by or characterized by protein deficiency. Therapeutic proteins encompass proteins, the delivery of which could ameliorate or correct such a deficiency. A therapeutic protein may act to replace a protein that is deficient in the disease or disorder. A therapeutic protein may be the protein that is deficient in the disease or disorder. However, a therapeutic protein need not necessarily be identical to the protein that is deficient in the disease or disorder. For instance, a therapeutic protein may be an active fragment or modified form of a deficient protein. A therapeutic protein may also partially or fully functionally compensate for the protein deficiency underlying the disease or disorder. A therapeutic protein may also ameliorate or correct downstream or secondary effects of the cellular deficiency in a particular protein.

**[00327]** In one embodiment, the therapeutic polypeptide comprises a cytotoxic polypeptide, preferably a polypeptide toxin or a functional fragment thereof. In one embodiment, the cytotoxic polypeptide comprises a catalytic domain from Diphtheria Toxin. In one embodiment, the cytotoxic polypeptide comprises a catalytic domain from a Chelona Toxin, such as from CT1 (SEQ ID NO: 2), CT2 (SEQ ID NO: 21), or CT3 (SEQ ID NO: 35) as described herein.

**[00328]** In on embodiment, the catalytic domain is from CT1. In one embodiment, the catalytic domain from the Chelona Toxin set forth in SEQ ID NO: 2 (CT1) has the amino acid sequence according to amino acid positions 1 to 186 of SEQ ID NO: 2.

**[00329]** In on embodiment, the catalytic domain is from CT2. In one embodiment, the catalytic domain from the Chelona Toxin set forth in SEQ ID NO: 21 (CT2) has the amino acid sequence according to amino acid positions 1 to 186 of SEQ ID NO: 21.

**[00330]** In on embodiment, the catalytic domain is from CT3. In one embodiment, the catalytic domain from the Chelona Toxin set forth in SEQ ID NO: 35 (CT3) has the amino acid sequence according to amino acid positions 1 to 191 of SEQ ID NO: 35.

**[00331]** In one embodiment, the therapeutic polypeptide comprises a protein that is deficient is a disease state, or a functional fragment thereof.

**[00332]** In one embodiment, the therapeutic polypeptide comprises Ras/Rap1-specific endopeptidase (RRSP) from *Vibrio vulnificus*, *e.g.*, as is set forth in SEQ ID NO: 18. In some embodiments the therapeutic polypeptide may be at least 80% identical to RRSP. In some embodiments the therapeutic polypeptide may be at least 90% identical to RRSP. In some embodiments the therapeutic polypeptide may be at least 95% identical to RRSP. In some

embodiments the therapeutic polypeptide may be at least 98% identical to RRSP. These sequence variant may retain substantially the same activity as full-length RRSP.

**[00333]** In one embodiment, the cargo molecule comprises an N-terminal cysteine residue for use in "click" chemistry bioconjugation.

**[00334]** In one embodiment, the cargo molecule comprises a nucleic acid molecule.

**[00335]** Percent sequence identifies described herein may be calculated across the full length of an alignment.

**[00336]** The amino acid sequences referred to herein may encompass sequence differences, in some embodiments compared to the references sequences (such as those set forth in Table 1, below). These may be variants, mutations, insertions, or deletions. In some applications, it may be important to ensure that the primary function of the protein is not substantially altered or abrogated, but this can be readily tested, e.g. using assays described herein. The amino acid sequences described herein may comprise a sequence of 75% or greater, 80% or greater, 85% or greater, 90% or greater, 95% or greater, or 99% or greater identity to the references sequences. The amino acid sequences may encompass conservative amino substitutions. Conservative amino acid substitutions which are known in the art are as follows with conservative substitutable candidate amino acids showing in parentheses: Ala (Gly, Ser); Arg (Gly, Gln); Asn (Gln; His); Asp (Glu); Cys (Ser); Gln (Asn, Lys); Glu (Asp); Gly (Ala, Pro); His (Asn; Gln); Ile (Leu; Val); Leu (Ile; Val); Lys (Arg; Gln); Met (Leu, Ile); Phe (Met, Leu, Tyr); Ser (Thr; Gly); Thr (Ser; Val); Trp (Tyr); Tyr (Trp; Phe); Val (Ile; Leu). Some so-called 'functional' variants, mutations, insertions, or deletions encompass sequences in which the function is substantially the same as that of the reference sequence, e.g. from which it is derived. This can be readily tested using assays similar to those described herein.

**[00337]** **<u>Nucleic Acids and Vectors</u>**

**[00338]** In one aspect, there is provided a nucleic acid encoding the recombinant polypeptide as defined here. In one embodiment, the nucleic acid is DNA or RNA. The RNA may be an mRNA.

**[00339]** A skilled person would readily appreciate there are many ways to encode the recombinant polypeptide described herein, e.g. due to degeneracy of the genetic code, all of which are encompassed in certain embodiments. Deletions, insertions, and substitutions may also be permitted if protein function remains substantially intact. For instance, nucleic acids may have 75% or greater, 80% or greater, 85% or greater, 90% or greater, 95% or greater, or 99% or greater identity to wild-type or references sequences may be encompassed. The above-noted nucleic acids could also be codon optimized depending on the organism or expression system in which it is intended to be expressed.

**[00340]** In one aspect, there is provided a vector comprising the nucleic acid as defined herein.

**[00341]** In one embodiment, there is provided a host cell comprising the nucleic acid as defined herein or the vector as defined herein. The host cell may be transformed or transfected.

**[00342]** **Compositions**

**[00343]** In one aspect, there is provided a composition comprising the recombinant polypeptide as defined herein, together with an acceptable excipient, diluent, or carrier.

**[00344]** In aspect embodiment, there is provided a pharmaceutical composition the recombinant polypeptide as defined herein, together with a pharmaceutically acceptable excipient, diluent, or carrier.

**[00345]** Pharmaceutically acceptable carriers include solvents, diluents, liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, or lubricants. Carriers may be selected to prolong dwell time for sustained release appropriate to the selected route of administration. Exemplary carriers include sugars such as glucose and sucrose, starches such as corn starch and potato starch, fibers such as cellulose and its derivatives, sodium carboxymethyl cellulose, ethyl cellulose, cellulose acetate, powdered tragacanth, malt, gelatin, talc, cocoa butter, suppository waxes, oils such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil, and soybean oil; glycols such as propylene glycol, esters such as ethyl oleate and ethyl laurate, agar, buffering agents such as magnesium hydroxide and aluminum hydroxide, alginic acid, pyrogen-free water, isotonic saline, Ringer's solution, ethyl alcohol, phosphate buffer solutions, non-toxic compatible lubricants such as sodium lauryl sulfate and magnesium stearate, coloring agents, releasing agents, coating agents, sweeteners, flavors, perfuming agents, preservatives, and antioxidants.

**[00346]** Compositions can be administered to subjects through any acceptable route, such as topically (as by powders, ointments, or drops), orally, rectally, mucosally, sublingually, parenterally, intracisternally, intravaginally, intraperitoneally, bucally, ocularly, or intranasally.

**[00347]** Liquid dosage forms for oral administration may include emulsions, microemulsions, solutions, suspensions, syrups and elixirs. Liquid dosage forms may contain inert diluents such as water or other solvents, solubilizing agents and emulsifiers such as ethyl alcohol, isopropyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils such as cottonseed, groundnut, corn, germ, olive, castor, and sesame oils, glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols and fatty acid esters of sorbitan, and mixtures thereof. Besides inert

diluents, the oral compositions can also include adjuvants such as wetting agents, emulsifying and suspending agents, sweetening, flavoring, and perfuming agents.

**[00348]** Dosage forms for topical or transdermal administration of an inventive pharmaceutical composition include ointments, pastes, creams, lotions, gels, powders, solutions, sprays, inhalants, or patches. The active agent is admixed under sterile conditions with a pharmaceutically acceptable carrier and any needed preservatives or buffers as may be required.

**[00349]** Injectable preparations, such as sterile injectable aqueous or oleaginous suspensions may be formulated according to the known art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile injectable solution, suspension or emulsion in a non-toxic parenterally acceptable diluent or solvent, for example, as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, U.S.P. and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid are used in the preparation of injectables. The injectable formulations can be sterilized prior to addition of spores, for example, by filtration through a bacterial-retaining filter, or by incorporating sterilizing agents in the form of sterile solid compositions which can be dissolved or dispersed in sterile water or other sterile injectable medium prior to use.

**[00350]** It is often desirable to slow the absorption of the agent from subcutaneous or intramuscular injection. Delayed absorption of a parenterally administered active agent may be accomplished by dissolving or suspending the agent in an oil vehicle. Injectable depot forms are made by forming microencapsule matrices of the agent in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of active agent to polymer and the nature of the particular polymer employed, the rate of active agent release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the agent in liposomes or microemulsions which are compatible with body tissues.

**[00351]** Compositions for rectal or vaginal administration are preferably suppositories which can be prepared by mixing the active agent(s) of this invention with suitable non-irritating excipients or carriers such as cocoa butter, polyethylene glycol or a suppository wax which are solid at ambient temperature but liquid at body temperature and therefore melt in the rectum or vaginal cavity and release the active agent(s).

**[00352]** Solid dosage forms for oral, mucosal or sublingual administration include capsules, tablets, pills, powders, and granules. In such solid dosage forms, the active agent

is mixed with at least one inert, pharmaceutically acceptable excipient or carrier such as sodium citrate or dicalcium phosphate, fillers or extenders such as starches, sucrose, glucose, mannitol, and silicic acid, binders such as, for example, carboxymethylcellulose, alginates, gelatin, polyvinylpyrrolidinone, sucrose, and acacia, humectants such as glycerol, disintegrating agents such as agar-agar, calcium carbonate, potato or tapioca starch, alginic acid, certain silicates, and sodium carbonate, solution retarding agents such as paraffin, absorption accelerators such as quaternary ammonium compounds, wetting agents such as, for example, cetyl alcohol and glycerol monostearate, absorbents such as kaolin and bentonite clay, and lubricants such as talc, calcium stearate, magnesium stearate, solid polyethylene glycols, sodium lauryl sulfate, and mixtures thereof.

**[00353]** Solid compositions of a similar type may also be employed as fillers in soft and hard-filled gelatin capsules using such excipients as milk sugar as well as high molecular weight polyethylene glycols and the like. The solid dosage forms of tablets, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings, release controlling coatings and other coatings well known in the pharmaceutical formulating art. In such solid dosage forms the active agent(s) may be admixed with at least one inert diluent such as sucrose or starch. Such dosage forms may also comprise, as is normal practice, additional substances other than inert diluents, such as tableting lubricants and other tableting aids such a magnesium stearate and microcrystalline cellulose. In the case of capsules, tablets and pills, the dosage forms may also comprise buffering agents. They may optionally contain opacifying agents and can also be of a composition that they release the active agent(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed manner. Examples of embedding compositions which can be used include polymeric substances and waxes.

**[00354]** The therapeutically effective amount may be determined on an individual basis or on the basis of the established amount necessary. The dosage for an individual subject is chosen in view of the subject to be treated. Dosage and administration may be adjusted to provide sufficient levels of the active agent(s) or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, contact with infectious agent in the past, potential future contact; age, weight, gender of the subject, diet, time and frequency of administration, drug combinations, reaction sensitivities, and tolerance/response to therapy. Sustained release compositions might be administered less frequently than fast-acting compositions.

**[00355]** **<u>Methods and Uses</u>**

**[00356]** In one aspect, there is provided a method of delivery a cargo molecule to a cell comprising contacting the cell with the recombinant polypeptide as defined herein.

**[00357]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for delivery of the cargo molecule to a cell.

**[00358]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for preparation of a medicament for delivery of the cargo molecule to a cell.

**[00359]** In one aspect, there is provided the recombinant polypeptide as defined herein for use in delivery of the cargo molecule to a cell.

**[00360]** In one aspect, there is provided a method treating cancer in a subject comprising administering to the subject the recombinant polypeptide as defined herein.

**[00361]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for treatment of cancer in a subject.

**[00362]** In one aspect, there is provided a use of the recombinant polypeptide as defined herein for preparation of a medicament for treatment of cancer in a subject.

**[00363]** In one aspect, there is provided the recombinant polypeptide as defined herein for use in treatment of cancer in a subject.

**[00364]** In one aspect, there is provided a method of alleviating enzyme or protein deficiency in a cell, comprising contacting a cell with the recombinant polypeptide as described herein.

**[00365]** In one aspect, there is provided a use of the recombinant polypeptide as described herein for alleviating enzyme or protein deficiency in a cell.

**[00366]** In one aspect, there is provided a use of the recombinant polypeptide as described herein for preparation of a medicament for alleviating enzyme or protein deficiency in a cell.

**[00367]** In one aspect, there is provided the recombinant polypeptide as described herein for use in alleviating enzyme or protein deficiency in a cell.

**[00368]** By 'alleviate, as used herein, is meant that the cargo molecule corrects or at least partially ameliorates the protein or enzyme deficiency, an aspect of the deficient protein or enzyme's function, or one or more of its downstream or secondary cellular effects or consequences.

**[00369]** In embodiments of the aforementioned methods and uses, the cargo may be released.

**[00370]** **EXAMPLES**

**[00371]** **Example 1**

**[00372]** **Introduction**

**[00373]** Engineered chimeric toxins has led to the emergence of novel therapeutics for challenging diseases, such as cancer. Immunotoxins are a class of biotherapeutics comprised of bacterial toxins, such as diphtheria toxin (DT), that have been repurposed into

cancer-targeted therapies – both by re-targeting their receptor binding domains (RBD) to target cancer receptors, and by delivering enzyme cargo that target intracellular oncoproteins. However, global vaccination programs against diphtheria have resulted in population-level immunity against DT, and DT-based therapeutics. To circumvent the issue of pre-existing neutralizing antibodies against DT, it was investigated whether distant homologs of DT sharing little sequence identity could retain the function of DT but avoid DT-specific neutralizing antibodies. Here, a putative gene sequence from *Austwickia chelonae* that is only 38% identical to DT has been structurally and functionally characterized. It has been named chelona toxin 1 (CT1). The x-ray crystallography structure thereof has been solved to 2.50Å and it was found that its structure was highly similar to DT. Using a variety of biochemical assays, a domain-by-domain analysis was undertaken to investigate the capacity of this DT-like protein to function as both a toxin and a new platform for therapeutic protein delivery. It has been demonstrated that while each domain of this novel protein can perform its respective function as a toxin, the translocase of CT1 (CT1-T) can be engineered to target non-native receptors and deliver non-native cargo into cells. Importantly, CT1 is not recognized by to pre-existing anti-DT antibodies found in human sera and is unexpectedly superior to DT at delivering cargo into cells. Chelona toxins provide novel insights into toxin biology and represents an improved platform for therapeutic protein delivery.

**[00374]** **Materials and Methods**

**[00375]** **Crystallization of the DT-like protein from *A. chelonae*.**

**[00376]** The closest DT-like protein from the species *Austwickia chelonae* (herein referred to as CT1 for "chelona toxin 1") (SEQ ID NO: 2) was chosen as the candidate for an alternative immunotoxin scaffold, due to the conservation of key residues identified to be important for DT functionality (Figure 5).

**[00377]** SEQ ID NO: 2 is derived from a combination of two ORFs (see GenBank Accession Nos. WP_143115263.1 and WP_040322835.1) representing two fragments of a toxin. When compared to the genomic sequence of *Austwickia chelonae* (see GenBank Accession No. NZ_BAGZ01000024.1), it appeared that a 1 base pair (bp) frameshift in the genomic sequence had caused a full-length toxin to be separated into the 2 ORFs. The reading frame was restored by deleting 1bp (NZ_BAGZ01000024.1 C41398), and the result was a full-length toxin was subsequently called "chelona toxin 1 (CT1)" (SEQ ID NO: 2). It is unclear whether the 1bp insertion was a sequence error or reflective of a genuine mutation in *Austwickia chelonae*. In any case, the 1bp insertion was removed to produce the protein and translocation domain use for the experiments described herein.

**[00378]** To determine the structure of CT1, the *E. coli* codon optimized gBlock gene fragment was ordered from Integrated DNA Technologies and cloned into the Champion™ pET SUMO *E. coli* expression system by Gibson Assembly.

**[00379]** A 50mL starter culture of NiCo21 (DE3) *E. coli* cells (New England Biolabs) were inoculated into 1L of LB medium and induced with 0.1mM IPTG at 18C for 18hours. Cells were centrifuged at 5000rpm and resuspended in lysis buffer (1% protease inhibitor cocktail, 1mg/mL lysozyme, 0.01% Pierce™ universal nuclease inhibitor, 20mM imidazole, 500mM NaCl, 20mM Tris-HCl pH 7.5). Cells were lysed with three passes through an Emulsiflex C3 (Avestin) at 15000 psi. Whole cell lysate was centrifuged at 18000xg and the supernatant was filtered through a 0.45um filter and passed over a HisTrap FF crude column (Cytiva). The protein was eluted with 50-75mM imidazole, buffer exchanged into 150mM NaCl, 20mM Tris-HCl pH 7.5, and incubated with SUMO protease overnight at 4C, to cleave the 6xHis-SUMO affinity tag. The protein was flowed over a HisTrap FF crude column and the flowthrough (protein) was collected and concentrated to 8 mg/mL by centrifugation.

**[00380]** Hanging drop vapour diffusion was used to grow crystals. The condition in which CT1 crystals were obtained contained 2uL of mother liquor (0.2M calcium chloride, 0.1M Tris-HCl pH 8.5, 25% (w/v) PEG4000) and 1uL of 8 mg/mL protein. The drop was dehydrated over 130uL of 2M $(NH_4)_2PO_4$ for 45 minutes prior to freezing in liquid nitrogen. Data was collected at the Advanced Photon Source on the 23-ID-D beamline.

**[00381]** Initial phases were determined using Phaser in the Phenix software package by using a multi-component search models with individual DT domains (C-domain residues 13-167, R-domain residues 391-531, T-domain residues 205-378) in which disordered loops had been removed. The structure was refined using iterative cycles of phenix.refine and autobuild.

**[00382]** **Protein Synthesis Assay**

**[00383]** Vero-nLucP cells (a nanoluciferase reporter strain of Vero cells) were plated at 5000 cells/well in 96-well white clear bottom plates (Corning). The following day, protein toxin was added and incubated for 24 hours, after which cells were read for luminescence signal using the NanoGlo® Luciferase Assay kit (Promega), on a SpectraMax M5e plate reader (Molecular Devices). Data was corrected to untreated cells (100% nanoluciferase signal).

**[00384]** **Liposomal dye release assay**

**[00385]** Unilamellar liposomes (DOPC, 0.8% DGS-NTA, Avanti Polar Lipids) were prepared as previously described. Briefly, 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) (Avanti Polar Lipids) was combined with 0.8% 1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt) (DGS-NTA[Ni]) (Avanti Polar Lipids), dried with $N_2$ and 1 hour in a vacuum dessicator. Lipids were resuspended in 20mM Tris pH

8, 35mM 8-Hydroxypyrene-1,3,6-trisulfonic acid (HPTS), 50mM p-xylene-bis-pyridinium bromide (DPX) (Thermo Fischer) and subject to 10X freeze-thaw cycles in dry ice and 42ºC water bath, and 10X extrusions using a 200um filter. The liposomes were then purified by gel filtration using a Hi Prep 16/60 Sephacryl S-300 HR column (GE Healthcare) and 150mM NaCl, 20mM Tris pH 8 buffer. Proteins were added in a ratio of 1:10,000 with liposomes, with a final liposome concentration of ~400uM, in 150mM citrate phosphate buffer ranging from pH 4.0 to 7.5, in 0.5 pH intervals. Assays were done in 96-well opaque plates (Corning), and fluorescence was monitored over a 20-minute interval, with readings being taken every 30 seconds (excitation 403nm, emission 510nm). Data were normalized to % of total HPTS fluorescence, by adding 0.3% Triton X-100.

**[00386] Serum Antibody Binding ELISA**

**[00387]** Nunc MaxiSorp™ plates (Thermo Fisher Scientific) were immobilized with 2000ng of protein after being blocked with 1% BSA, and were subsequently incubated with human serum (Pooled Human Serum frozen, Cedarlane) at various dilutions, for 1 hour. Wells were washed with PBST (0.01% tween) and then incubated with an anti-human IgG antibody conjugated to HRP (Abcam, ab102420), that was developed using TMB reagent (Thermo Fisher Scientific). Absorbance was read at 630nm and protein wells were corrected to control wells (no-protein, +human serum).

**[00388] Serum Toxicity Assays**

**[00389]** Protein toxins were incubated with either human serum (Pooled Human Serum frozen, Cedarlane) or mouse serum (Mouse serum sterile frozen, Cedarlane) in a 1:1 ratio, for 30 minutes at room temperature. Sample was then added to Vero-nLucP cells that had been plated to 5000 cells/well the previous day, in a 96-well white clear bottom plate (Corning). Cells were incubated for 24hours, upon which cells were lysed and assessed for luminescence signal. Values were corrected to serum only treated cells, which represented 100% nanoluciferase signal and 100% protein synthesis.

**[00390] <u>Results</u>**

**[00391] Structural characterization of DT-like protein from *A. chelonae***

**[00392]** With the goal of finding an alternative DT-like immunotoxin platform unsusceptible to pre-existing anti-DT antibodies yet functionally active, the evolutionarily closest DT-like protein outside the *Corynebacterium* genus was chosen. By sequence, the DT-like protein from *A. chelonae* (CT1) is 38% identical to DT. The catalytically active residue (E148) and residues important for substrate (NAD) binding and coordination (H21, Y54, Y65) are all conserved, as is the furin recognition site, and the disulfide bond formed between C186 and C201 (in DT) is also present in the CT1 sequence (Table 1). Furthermore, key histidine residues and charged residues involved in the pH-dependent

unfolding and pore formation of the translocation domain were found to be conserved in the predicted translocation domain of the sequence from CT1 (Table 1). To test whether the putative domains of the CT1 were functional, the protein was cloned, expressed and purified to study structurally and functionally (Figure 1B).

**[00393]** Hanging drop vapour diffusion was used to obtain an x-ray crystallography structure of CT1. The protein successfully crystallized and diffracted to 2.50Å. It was not possible to use the full-length DT structure (pdb 1MDT) as a search model for molecular replacement. However, using partial search models with 1MDT, the structure was solved (Figure 1C). CT1had a RMSD of 2.4Å to DT, and retained the same three domain, Y-shaped architecture as DT.

**[00394]** **The C-domain of CT1 is functional and has the same intracellular target as DT**

**[00395]** Structural alignments of CT1-C (the catalytic domain of CT1) to $DT_C$ (the DT catalytic domain) show good structural conservation between key residues required for $DT_C$ functionality. In order to test whether CT1-C was functional, a chimera was generated in which $DT_C$ was replaced with CT1-C (referred to as CT1-C -$DT_T$-$DT_R$). The chimeric protein was tested on HEK293T cells with and without a gene knockout of DPH4 ($DPH4^{-/-}$). These cells are defective in the diphthamide synthesis pathway, and produce eukaryotic elongation factor 2 (eEF-2) without a diphthamide modification, and are therefore completely resistant to DT (Figure 2A and 2B). Wildtype HEK293T cells were susceptible to both wildtype DT and CT1-C -$DT_T$-$DT_R$, while the diphthamide knockout cells were not. This indicates that the catalytic domain of CT1 has the same intracellular target as DT.

**[00396]** **The C-domain of CT1 is efficiently released**

**[00397]** An important part of DT's intoxication mechanism is the release of the C-domain from the rest of the molecule, upon entry into the cytosol. The furin protease recognition site is conserved in CT1 (**R**AK**R** in CT1). To confirm that the furin site is recognized and the C-domain is released from the rest of the molecule, DT and CT1 were incubated with mammalian cell lysate overnight at 37C. Both DT and CT1 were cleaved between the C and T domains (Figure 2C). When DTC-CT1-F-DTT-DTR was tested on Vero cells, it was similarly toxic to DT (Figure 3A), suggesting it is efficiently cleaved by furin.

**[00398]** **CT1 contains a functional translocase**

**[00399]** Endosomal acidification leads to the refolding and insertion of $DT_T$ into the endosomal membrane, and subsequent translocation of the C-domain into the cytosol. This process is thought to be initiated and driven by nine charged residues in $DT_T$, of which six are conserved in CT1-T. To test whether CT1-T forms DT-like pores, the isolated T-domain was purified and tested *in vitro* for its capacity to release dye from liposomes (data not

shown). CT1-T showed a pH-dependent increase in dye release, with the onset of dye release (interpreted as pore formation) at pH 5.5 (similar to $DT_T$).

**[00400]** It was further investigated whether the translocase deliver cargo into cells. To this end, a chimeric DT was generated in which $DT_T$ was swapped for CT1-T ($DT_C$- CT1-T - $DT_R$) and measured its effect on protein synthesis. CT1-T successfully delivered $DT_C$ into cells, as shown by the decrease in signal observed with increasing amounts of chimeric toxin (Figure 3B. Furthermore, when $DT_C$ was swapped for a non-native cargo (Ras/Rap1 Specific Protease; RRSP) targeting Ras oncoproteins, it was shown to successfully deliver this structurally distinct cargo (as shown by cell viability on CFPAC-1 cells, and western blot assessing intracellular Ras levels; Figure 3C and 3D). Taken together, these experiments demonstrated that CT1-T is a functional translocase capable of delivering various cargo into the cytosolic component of cells.

**[00401]** **The translocation domain of CT1 can accommodate diverse receptor-binding domains**

**[00402]** Having demonstrated that CT1-T could tolerate manipulation on the N-terminus (can translocate $DT_C$ and RRSP), it was next assessed whether CT1-T could tolerate such manipulation on the C-terminus. To this end, a chimera was generated in which the receptor binding domains ($DT_R$) of $DT_C$- CT1-T -$DT_R$ and RRSP- CT1-T -$DT_R$ were swapped with a Her3 (human epidermal receptor 3) targeting affibody (ZHer3:08699, referred to hereafter as ZHer3) and tested these constructs on HPAF II cells (Figure 3E and 3F). It was found that not only could CT1-T be re-targeted to Her3 expressing cells, but surprisingly, that CT1-T -containing constructs appeared more efficient than $DT_T$ at cargo delivery.

**[00403]** **CT1 is not recognized or neutralized by human serum**

**[00404]** To quantify the level of pre-existing anti-DT or anti-CT1 antibodies in human serum, an ELISA assay was used in which either DT or CT1 was immobilized on high-bind plates and incubated with varying amounts of pooled human sera, after which an anti-IgG antibody conjugated to HRP was used to determine levels of antibody binding (Figure 4A). While DT showed a dose-dependent increase in signal in the presence of human sera, CT1 did not, indicating that CT1 is not recognized by antibodies in human serum.

**[00405]** To confirm that DT is neutralized by antibodies in human sera, DT was incubated for 30 minutes with pooled human sera and then added it to Vero cells. A 6-log shift in toxicity of DT was observed, indicating neutralization by human sera (Figure 4B). In contrast, CT1 was not neutralized by human sera. This indicates that CT1 is not neutralized by anti-DT antibodies. This provides CT1-based delivery vectors an advantage over DT-based vectors by circumventing the issue of pre-existing neutralizing anti-DT antibodies.

**[00406]** In addition to the translocation domain for CT1 (SEQ ID NO: 3), the translocation domains of SEQ ID NOs: 12 and 14 have also been established to be functional.

**[00407]** On the basis of the results obtained, the related polypeptide sequences from other bacterial strains and species described herein, and as set forth in SEQ ID NOs: 4 to 11, 13, 15 to 17, and 36 to 48 (and their related sequences), are also expected to be functional translocation domains that are active within polypeptide constructs as described herein. Figure 9 provides a phylogenetic tree showing relationships between translocase domains.

**[00408]** **Example 2**

**[00409]** To evaluate the function of the novel translocases relative to DT, each translocase sequence, "T", was cloned between the intracellular RAS cleaving enzyme RRSP (Ras/Rap1 Specific Peptidase, SEQ ID NO: 18; viz. the cargo) and a dual receptor binding domain known as ZHer3-A20 (consisting of an affibody against Her3, SEQ ID NO: 19, and a peptide against $\alpha v\beta 6$ integrin known as A20FMDV2 (A20), SEQ ID NO: 20) yielding the construct RRSP-T-ZHer3-A20 (where "T" indicates the translocase). A range of protein concentrations of each identified construct was incubated with human pancreatic adenocarcinoma (HPAF-II) cells for 72-hours (Fig. 6A), and/or with epithelial-like cell (H358 cells) for 72-hours (Fig. 6B). As RRSP induces apoptosis after being delivered into cells, translocation was quantified by measuring the ability of each construct to kill human pancreatic adenocarcinoma cells after a 72-hour incubation.

**[00410]** Figures 6A and 6B show results of functional characterization of translocases. Using the RRSP-T-ZHer3-A20 template, the function of each translocase was quantified. The concentration of a given construct that resulted in reduction of cell viability by 50% of maximal toxicity (EC50) is represented in the bar graph. A lower value represents a more efficient translocase. The translocase from *S. pinicola* was not expressed and so could not be evaluated. The translocases from *S. piniterrae* and *L. tulafanense* were determined to be non-functional in the assay as no toxicity was observed up to the highest dose tested (100nM). The remaining five translocases were functional in the assay. The translocases of the proteins from *A. chelonae* (SEQ ID NO: 2) and *A. chelonae LK16-18* (SEQ ID NO: 17) were the most active on cells and better than DT's translocase. Activity levels observed indicate that the RRSP is efficiently released.

**[00411]** The translocase domains *Streptomyces albireticuli* and *Seinonella peptonophila* have also been shown to be functional (see Sugiman-Marangos et al. 2022, which is incorporated by reference in its entirety).

**[00412]** The diphtheria toxoid vaccine is part of global vaccination programs that serve to protect against the disease diphtheria. Anti-DT antibodies in human sera prevent the

actions of DT by binding to DT and neutralizing its function. Unfortunately, these same antibodies also bind to and neutralize DT-based therapeutics. An ELISA was performed to evaluate the degree to which anti-DT antibodies in human sera recognize full-length toxins DT and CT1 (SEQ ID NO: 2), and the translocases DT-T and CT1-T (SEQ ID NO: 3). As shown in the ELISA data in Figure 7A, antibodies in human sera recognize DT, but show no binding to CT1. Similarly, high titres are seen against the Translocase from DT, but not CT1 (Figure 7B).

**[00413]** Figure 7A and 7B together show that pre-existing anti-DT antibodies in human serum do not bind or neutralize CT1-based immunotoxins. The respective protein was immobilized on plates and incubated with varying amounts of pediatric human serum sample, after which an anti-IgG antibody conjugated to HRP was used to determine levels of antibody binding. DT and DT-T both showed a dose dependent increase in signal with human serum, while CT1 and CT1-T did not. This indicates that full length CT1 and just its translocase do not contain epitopes recognized by antibodies in human serum.

**[00414]** Next, to evaluate the degree to which pre-existing anti-DT antibodies in human sera neutralize the function of DT-based therapeutics as well as the corresponding therapeutics based on novel translocation domains, DT- and the CT-based immunotoxins from A. chelonae LK16-18 were incubated with human sera. DT- and CT-based immunotoxins were cloned and purified, where the C- and T-domains of the respective toxin were recombinantly attached to ZHer3-A20 (either $DT_{1-389}$-ZHer3-A20 or $CT2_{1-391}$-ZHer3-A20). As shown in Figure 8, in the absence of human sera (PBS), DT, $DT_{1-389}$-ZHer-3-A20 and $CT2_{1-391}$-ZHer3-A20 were all toxic to human cells. In the presence of human sera (human), a rightward shift in the toxicity curve is seen for DT and $DT_{1-389}$-ZHer3-A20, indicating the human sera inhibited/neutralized their function. Conversely, human sera had no effect on the toxicity curve for $CT2_{1-391}$-ZHer3-A20, indicating that human sera did not recognize or inhibit/neutralize CT-based therapeutics.

**[00415]** Figure 8 shows that anti-DT antibodies do not neutralize CT-based immunotoxins. Immunotoxins were incubated for 30 minutes with either a PBS control or human serum, and then added to cells. Cell viability was assessed at 72 hours, and it was found that while $DT_{1-389}$-ZHer3-A20 had ~3-log decrease in toxicity upon human serum incubation, $CT2_{1-391}$-ZHer3-A20 had no shift.

**[00416]** **Example 3**

**[00417]** **Example Annotated Construct Sequences**

**[00418]** RRSP – T (*C. diphtheriae*) - ZHer3-A20 (SEQ ID NO: 22)

**[00419]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF

MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI
DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCINLDWDVIR
DKTKTKIESLKEHGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVF
AGANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALS
SLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPFGGGGSGGG
GSAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKG
GGGSGGGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00420]** In the above:

**[00421]** residues 1-510 = RRSP

**[00422]** residues 511-724 = T domain from *C. diphtheriae*

**[00423]** residues 725-734 = G4S2 linker

**[00424]** residues 735-792 = ZHer3:08699 affibody

**[00425]** residues 793-802 = G4S2 linker

**[00426]** residues 803-822 = A20FMDV2 peptide

**[00427]** residues 823-840 = thrombin cleavage site and strep-tag II

**[00428]** RRSP – T (*A. chelonae*) - ZHer3-A20 (SEQ ID NO: 23)

**[00429]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI
AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF
MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI
DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCLSKINWKNV
REKADALTKKVHADKEFMDKLSTHHQRGEAPSVEKTTALHNALLEHESFSALKGARASGKV
GAAASTAAWGVAVAQAFTDPKADALTKTAATLSVVPGLGQALGIADGIKHENTEEIVVQSISL
AGLLAAQAIPVVGEAVDFGLLVYQLVETIVDLATHLSSAAANPAYSPGHKTQPFGGGGSGG
GGSAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPK
GGGGSGGGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00430]** In the above:

**[00431]** residues 1-510 = RRSP

**[00432]** residues 511-520 = G4S2 linker

**[00433]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00434]** residues 537-711 = T domain from *A. chelonae*

**[00435]** residues 712-723 = linker sequence from *C. diphtheriae*

**[00436]** residues 724-733 = G4S2 linker

**[00437]** residues 734-791 = ZHer3:08699 affibody

**[00438]** residues 792-801 = G4S2 linker

**[00439]** residues 802-821 = A20FMDV2 peptide

**[00440]** residues 822-839 = thrombin cleavage site and strep-tag II

**[00441]** RRSP – T (*A. chelonae LK16-18*) - ZHer3-A20 (SEQ ID NO: 24)

**[00442]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCLSKVKWEQ VREKSKKIIDNVKDNPEFMKKLSAHHERGSAPTTEKITALHNELLDHESFSALKGARSSAGTA ATAASAAAWGLAVAQAFTNPKADDLTKATAVLSAVPGLGQALGIADGIKHHNTEEIVVQSISL TALIAAQAIPVVGELVDFGLLAYQLVESIIDLTRQLSVITANPAYSPGHKTQPFGGGGSGGGG SAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGG GGSGGGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00443]** In the above:

**[00444]** residues 1-510 = RRSP

**[00445]** residues 511-520 = G4S2 linker

**[00446]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00447]** residues 537-712 = T domain from *A. chelonae* LK16-18

**[00448]** residues 713-724 = linker sequence from *C. diphtheriae*

**[00449]** residues 725-734 = G4S2 linker

**[00450]** residues 735-792 = ZHer3:08699 affibody

**[00451]** residues 793-802 = G4S2 linker

**[00452]** residues 803-822 = A20FMDV2 peptide

**[00453]** residues 823-840 = thrombin cleavage site and strep-tag II

**[00454]** RRSP – T (*A. TVS 96-490-7B*) - ZHer3-A20 (SEQ ID NO: 25)

**[00455]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI
AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF
MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI
DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCLSNIKWERV
RQKSEELIKKLKDDENVKKAVQERKEGVKPTSDDLQNLHKALIDHESFKELKSVHSNGVKA
MDAANAALWGANVARVFSDSKSDGLEKATAALAAVPGLGQVMGVADGVTHHNTEEVVVQ
SVALAGFIAAQAIPVVGEIVDIGVLAYQFVEGVIDLSRQMMTSNARPAYSPGHKTQPFGGGG
SGGGGSAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQ
APKGGGGSGGGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00456]** In the above:

**[00457]** residues 1-510 = RRSP

**[00458]** residues 511-520 = G4S2 linker

**[00459]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00460]** residues 537-711 = T domain from A. *TVS 96-490-7B*

**[00461]** residues 712-723 = linker sequence from *C. diphtheriae*

**[00462]** residues 724-733 = G4S2 linker

**[00463]** residues 734-791 = ZHer3:08699 affibody

**[00464]** residues 792-801 = G4S2 linker

**[00465]** residues 802-821 = A20FMDV2 peptide

**[00466]** residues 822-839 = thrombin cleavage site and strep-tag II

**[00467]** RRSP – T (*S. klenkii*) - ZHer3-A20 (SEQ ID NO: 26)

**[00468]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI
AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF
MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI

DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCLPVDWDKV
EEKAKATAKKVAQDAEHVEKLPKRSPKGPTWGEAHSTAELSHAKVREVSGAHVAASAAGV
GTWVYGMAKTFSDKDATTLDKVAVTGAVVPGLGQALGIADGIQHGDPEAIAVNAVALAALA
AAQVVPVVGEVVDAVLLTEQLVEVLVDVFRTATADPAYSPGHKTQPFGGGGSGGGGSAEA
KYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSG
GGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00469]** In the above:

**[00470]** residues 1-510 = RRSP

**[00471]** residues 511-520 = G4S2 linker

**[00472]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00473]** residues 537-700 = T domain from *S. klenkii*

**[00474]** residues 701-712 = linker sequence from *C. diphtheriae*

**[00475]** residues 713-722 = G4S2 linker

**[00476]** residues 723-780 = ZHer3:08699 affibody

**[00477]** residues 781-790 = G4S2 linker

**[00478]** residues 791-810 = A20FMDV2 peptide

**[00479]** residues 811-828 = thrombin cleavage site and strep-tag II

**[00480]** RRSP – T (*S. sp TLI053*) - ZHer3-A20 (SEQ ID NO: 27)

**[00481]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI
AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF
MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI
DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCIPAASWEAI
EHRSKEMASAVARDTEYTKSLPNRHPKGPTWSEAHTTSTSTHARVSAKSGAHVAVGAFAV
GSWIYGMSETFANKNVTTLDKAAATVAIVPGIGHALGIAAALEHHDIEGVVVNAISIAALAAAQ
VVPVVGEIVDAALLAEQLVEVLVHVFRASTTEPAYSPGHKTQPFGGGGSGGGGSAEAKYAK
EKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGGG
SNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00482]** In the above:

**[00483]** residues 1-510 = RRSP

**[00484]** residues 511-520 = G4S2 linker

**[00485]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00486]** residues 537-701 = T domain from *S. sp TLI053*

**[00487]** residues 702-713 = linker sequence from *C. diphtheriae*

**[00488]** residues 714-723 = G4S2 linker

**[00489]** residues 724-781 = ZHer3:08699 affibody

**[00490]** residues 782-791 = G4S2 linker

**[00491]** residues 792-811 = A20FMDV2 peptide

**[00492]** residues 812-829 = thrombin cleavage site and strep-tag II

**[00493]** RRSP – T (*L. tulufanense*) - ZHer3-A20 (SEQ ID NO: 28)

**[00494]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI
AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF
MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI
DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCERKTLDKLD
RKKIETRAKQVISKLAQDGDLRRELPRRTTTGHAHEEVRGVLAKSRGALHEIHGAATTVAMP
LATVAWVEDMARVFREKNATTLDKAATVSEIVPVAGQVLGMADGIAHRDAETVAVNAVVLA
AIAVSQAVPVVGELVDLGLTAYAVVDVVVRLFGPAREVPITQESYWPAYSPGHKTQPFGGG
GSGGGGSAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDS
QAPKGGGGSGGGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00495]** In the above:

**[00496]** residues 1-510 = RRSP

**[00497]** residues 511-520 = G4S2 linker

**[00498]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00499]** residues 537-714 = T domain from *L. tulufanense*

**[00500]** residues 715-726 = linker sequence from *C. diphtheriae*

**[00501]** residues 727-736 = G4S2 linker

**[00502]** residues 737-794 = ZHer3:08699 affibody

**[00503]** residues 795-804 = G4S2 linker

**[00504]** residues 805-824 = A20FMDV2 peptide

**[00505]** residues 825-842 = thrombin cleavage site and strep-tag II

**[00506]** RRSP – T (*S. piniterrae*) - ZHer3-A20 (SEQ ID NO: 29)

**[00507]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI
AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF
MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI
DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCELAKDKAR
DILSEAGNEVSLPQRDSDGLSKQEITATAEATRSKLGTGVHGAVSAAMVADWAHDVARTFA
DPKATKLDKAAAVTAIAPVIGQAVNIADGIQHHDKKTIVVNSLVLAAVVAAQAVPAVGEVVDA
AIVADFVVEKLVGWFTPTAKPGPEHVAPAYSPGHKTQPFGGGGSGGGGSAEAKYAKEKYN
AYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGGGSNAV
PNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00508]** In the above:

**[00509]** residues 1-510 = RRSP

**[00510]** residues 511-520 = G4S2 linker

**[00511]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00512]** residues 537-695 = T domain from *S. piniterrae*

**[00513]** residues 696-707 = linker sequence from *C. diphtheriae*

**[00514]** residues 708-717 = G4S2 linker

**[00515]** residues 718-775 = ZHer3:08699 affibody

**[00516]** residues 776-786 = G4S2 linker

**[00517]** residues 786-805 = A20FMDV2 peptide

**[00518]** residues 806-823 = thrombin cleavage site and strep-tag II

**[00519]** RRSP – T (*S. pinicola*) - ZHer3-A20 (SEQ ID NO: 30)

**[00520]** GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQLDLVHQAKGRDQI
AASFELNKKINDYIAEHPTSGRNQALTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF
MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLGFDAKYHVDLGEHYADFK
QWLETSQSNGLLSKATLDESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISG
DSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDAWDRRYSGTGYDELTNKLASA
TGVDEQLAVLLDDRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQPLI

DRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIVALDANSSARPNVQGTEHGLMYRA
GAANNIAVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSNQFTVE
QDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGSCAGNRVRRSVGSSLSCKVNWKDAY
DRSMNIAQDIDGSAEFRATAPARPAAGKAMPEADVTRLVSTSEDFLEKSTKTSGVKKALKAI
NNQKMISWGSLLSNALANSHTWSDKNATNLDKAYAVVGGVPVLGEVIGIASGIDKQDAESIA
VNTLSLVGIVAATVCPPLGATVEFVMIGYTAIKLMLSWFTVETIPAYSPGHKTQPFGGGGSG
GGGSAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAP
KGGGGSGGGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK

**[00521]** In the above:

**[00522]** residues 1-510 = RRSP

**[00523]** residues 511-520 = G4S2 linker

**[00524]** residues 521-536 = *C. diphtheriae* sequence with furin protease recognition site

**[00525]** residues 537-712 = T domain from *S. pinicola*

**[00526]** residues 713-724 = linker sequence from *C. diphtheriae*

**[00527]** residues 725-734 = G4S2 linker

**[00528]** residues 735-792 = ZHer3:08699 affibody

**[00529]** residues 793-802 = G4S2 linker

**[00530]** residues 803-822 = A20FMDV2 peptide

**[00531]** residues 823-840 = thrombin cleavage site and strep-tag II

**[00532]** <u>DT-Her3-A20 (SEQ ID NO: 31)</u>

**[00533]** GADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDW
KGFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTE
PLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGKRG
QDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGPIKNKMSESPN
KTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAAWAVNVAQVIDSETADN
LEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFV
ESIINLFQVVHNSYNRPAYSPGHKTQPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQLPNLT
KYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGGGSNAVPNLRGDLQVLAQ
KVARTRQALVPRGSAWSHPQFEK

**[00534]** In the above:

**[00535]** residues 1-389 = sequence from *C. diphtheriae*

**[00536]** residues 390-399 = G4S2 linker

**[00537]** residues 400-457 = Her3:08699 affibody

**[00538]** residues 458-467 = G4S2 linker

**[00539]** residues 468-487 = A20FMDV2 peptide

**[00540]** residues 487-505 = thrombin cleavage site and strep-tag II

**[00541]** CT2-ZHer3-A20 (SEQ ID NO: 32)

**[00542]** YANDAVIADQSKTVDSFTSYHGAKPESFESVLAGIKKPESGSQGNHDPEWK GFYTTDNKHAAAGYTVSDESVMTGKAGGVVKVTYPGKTRVLAVKPLSATELKTYLGLAADK PLIDQLNNKDFINKFGEGASRVVLQMPFADGTSDVEYIHNWEDATQLQVATEVRFDNLGKR GQDEMNRYMNLANCPSVSAVRVKRNPAKLCLSKVKWEQVREKSKKIIDNVKDNPEFMKKL SAHHERGSAPTTEKITALHNELLDHESFSALKGARSSAGTAATAASAAAWGLAVAQAFTNP KADDLTKATAVLSAVPGLGQALGIADGIKHHNTEEIVVQSISLTALIAAQAIPVVGELVDFGLLA YQLVESIIDLTRQLSVITANPPTEVTHSSKLAGGGGSGGGGSAEAKYAKEKYNAYYEIWQLP NLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGGGSNAVPNLRGDLQV LAQKVARTRQALVPRGSAWSHPQFEK

**[00543]** In the above:

**[00544]** residues 1-391 = sequence from *A. chelonae LK16-18*

**[00545]** residues 392-401 = G4S2 linker

**[00546]** residues 402-459 = Her3:08699 affibody

**[00547]** residues 460-469 = G4S2 linker

**[00548]** residues 470-489 = A20FMDV2 peptide

**[00549]** residues 491-507 = thrombin cleavage site and strep-tag II.

**Table 1: Table of Sequences**

| Full-length Proteins | | | |
|---|---|---|---|
| **SEQ ID NO:** | **Species** | **Sequence** | **GenBank Accession** |
| 1 | *Corynebacterium diphtheriae* | GADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQKPKS GTQGNYDDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGV VKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGT EEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALS VELEINFETRGKRGQDAMYEYMAQACAGNRVRRSVGSSLS CINLDWDVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSE EKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAA WAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGA VHHNTEEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNF VESIINLFQVVHNSYNRPAYSPGHKTQPFLHDGYAVSWNT VEDSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGK LDVNKSKTHISVNGRKIRMRCRAIDGDVTFCRPKSPVYVG NGVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDH TKVNSKLSLFFEIKS | From WP_072564851.1 |
| 2 | *Austwickia chelonae* (CT1) | YANDVVLKDQSTTVDSFTSYHGAKPESFNAVLTGIKKPEK GSQGNNDPDWKGFYTTDNKHAAAGYTVSDESVLSGKAGGV VRVTYPGKTRILAVKSLSAAELKGKLGLDSAKPLIDQLND KSFLEKYGDGANRVVLKMPFADGTEDSEFIHNWKDAEQLS VETEVRFDNLGKRGQDAMNSYMNMANCPSTSAVRAKRSPG KICLSKINWKNVREKADALTKKVHADKEFMDKLSTHHQRG EAPSVEKTTALHNALLEHESFSALKGARASGKVGAAASTA AWGVAVAQAFTDPKADALTKTAATLSVVPGLGQALGIADG IKHENTEEIVVQSISLAGLLAAQAIPVVGEAVDFGLLVYQ LVETIVDLATHLSSAAANPPTEATDSVRPAVSLGLRAGWK TEEDAKLHIGSPYGMKFQRIVLSAEEGKEIPFVRAAVAVD SKFLKINGPRSFVVQNGIKTPMACFETEGNLAFCRPSRPI FLSSSSPATLHLSYVTNEHENGTIKNPTVDILGQRIVENK VITANKVSLVYKVDSSNTL | N/A |
| 21 | *Austwickia chelonae LK16-18* (CT2) | YANDAVIADQSKTVDSFTSYHGAKPESFESVLAGIKKPES GSQGNHDPEWKGFYTTDNKHAAAGYTVSDESVMTGKAGGV VKVTYPGKTRVLAVKPLSATELKTYLGLAADKPLIDQLNN KDFINKFGEGASRVVLQMPFADGTSDVEYIHNWEDATQLQ VATEVRFDNLGKRGQDEMNRYMNLANCPSVSAVRVKRNPA KLCLSKVKWEQVREKSKKIIDNVKDNPEFMKKLSAHHERG SAPTTEKITALHNELLDHESFSALKGARSSAGTAATAASA AAWGLAVAQAFTNPKADDLTKATAVLSAVPGLGQALGIAD GIKHHNTEEIVVQSISLTALIAAQAIPVVGELVDFGLLAY QLVESIIDLTRQLSVITANPPTEVTHSSKLAKSNGLLAGW KTDQDGVLSLGAPHGMKTQLISMSAEKGQEIPFTGAMIAV | WP_162873017.1 Y64-L603 |

| | | | |
|---|---|---|---|
| | | KKDLLTVNTPRVFVVQNGIRIPLRCSAAGETLTFCRPNHP VWLSEQHQATLHLSYVTQENESEKIDNPTVDILGQKIVDN KVSTANKVSLVYEVDRSNKL | |
| 35 | *Austwickia sp. TVS 96-490-7B* (CT3) | GAFASPISPAIDKEKSKVVENFTSYHGTKHQYIENVLKGI NKPETGSQGNHDPDWKGFYTTDSKYAAAGYTPDEENPMSG KAGGVVKVSYTGKTKILALKDMPAGKIKGHLGLDPNKPLT EQIKDQEFINKYGEGADRVVLQMPFAEGSSEVEYINNWDN AKDLKVEPEVRFDQHGKRGQDAMYEYMALASCESAGKAKR SVSKTCLSNIKWERVRQKSEELIKKLKDDENVKKAVQERK EGVKPTSDDLQNLHKALIDHESFKELKSVHSNGVKAMDAA NAALWGANVARVFSDSKSDGLEKATAALAAVPGLGQVMGV ADGVTHHNTEEVVVQSVALAGFIAAQAIPVVGEIVDIGVL AYQFVEGVIDLSRQMMTSNARGIAEKRQVVDIKDANGLSG GWLTPQDSKIHLPPHYGTKTQRLAVSANGDHNLPITGFKT SINPKFFNINESGSFIVQNGLKVPMNCSEEASKTEAVCIP VHPIWLTDKNNAIIHVSYQTKSPEGSRIDNPRLEVWGQIM HTEEGPKTNTINLGYSVDQEMPCYSSIKFENGGDYVGRLA VNSKDGKVTHSKMSGMRDKDKVDLSAFNSGEVLTTEISPT GGGAPTPGPSIKYCSDGGIEASIKSWGTKDKAFLGFN | WP_219106995.1 G27-N663 |
| **Translocation Domains** | | | |
| **SEQ ID NO:** | **Species** | **Translocation Domain Sequence** | **Source Sequence or GenBank Accession (full-length sequence with coordinates)** |
| 3 | *Austwickia chelonae (*CT1-T*)* | LSKINWKNVREKADALTKKVHADKEFMDKLSTHHQRGEAP SVEKTTALHNALLEHESFSALKGARASGKVGAAASTAAWG VAVAQAFTDPKADALTKTAATLSVVPGLGQALGIADGIKH ENTEEIVVQSISLAGLLAAQAIPVVGEAVDFGLLVYQLVE TIVDLATHLSSAAAN | SEQ ID NO: 2 L204-N378 |
| 4 | *Streptosporangium nondiastaticum* | LPVDWDKVEEKAKATAKKVAQDAEHVERLPKRSPKGPTWG EAHSTAELTHAKVREVSGAHVAASAAGVGTWVYGMAKTFS DKNATTLDKVAVTGAVVPGLGQALGIADGIQHGDPEAIAV NAVALAALAAAQVVPVVGEVVDAVLLTEQLVEVLVDVFRT ATAD | PSJ28985.1 L231-D394 |
| 5 | *Streptomyces sp.TLI 053* | IPAASWEAIEHRSKEMASAVARDTEYTKSLPNRHPKGPTW SEAHTTSTSTHARVSAKSGAHVAVGAFAVGSWIYGMSETF ANKNVTTLDKAAATVAIVPGIGHALGIAAALEHHDIEGVV VNAISIAALAAAQVVPVVGEIVDAALLAEQLVEVLVHVFR ASTTE | SDT83331.1 I198-E362 |

| | | | |
|---|---|---|---|
| 6 | *Streptomyces sp. SLBN-118* | LSEMDGEKVTKGAQSVADTLKGHGNLQPLVKRTRAEPRTQ QYLPPTRQQIDDAISAAQRVHPGVDTAMKHLGSAQALAGF ALDVSQGEPLRAAADVTAMLPRIGPGLGNALGLADSIVHD DPEGIAVNAIGLGAFIVAQAIPVIGELADLGFAAYALIET LKSMFAGPTLG | WP_160159328.1 L631-G801 |
| 7 | *Streptomyces sp. AA8* | LPVDWDKVEEKAKATAKKVAQDAEHVERLPKRNPKGPTWG EAHSTAELTHAKVRAVSGAHVAATAAGVGTWVYGMAKTFS DKDATTLDKVAVTGAVVPGLGQALGIADGIQHGDPEAVAV NAVALAALAAAQVVPVVGEVVDAALLTEQLVEVLVDVFRT ATAD | WP_168096531.1 L241-D404 |
| 8 | *Streptomyces roseoverticillatus* | LPVDWDKVEEKAKATAKKVAQDAEHVERLPKRNPKGPTWG EAHSTAELTHAKVRAVSGAHVAATAAGVGTWVYGMAKTFS DKDATTLDKVAVTGAVVPGLGQALGIADGIQHGDPEAVAV NAVALAALAAAQVVPVVGEVVDAALLTEQLVEVLVDVFRT ATAD | WP_078659863.1 L231-D394 |
| 9 | *Streptomyces piniterrae* | ELAKDKARDILSEAGNEVSLPQRDSDGLSKQEITATAEAT RSKLGTGVHGAVSAAMVADWAHDVARTFADPKATKLDKAA AVTAIAPVIGQAVNIADGIQHHDKKTIVVNSLVLAAVVAA QAVPAVGEVVDAAIVADFVVEKLVGWFTPTAKPGPEHVA | JZ58907.1 E225-A383 |
| 10 | *Streptomyces MBT76* | LPVDWDKVEEKAKATAKKVAQDAEHAERLPKRNPKGPTWG EAHSTAELTHAKVREVSGTHVAATAAGVGTWVYGMAKTFS DKDATTLDKVAVTGAVVPGLGQALGIADGIQHGDPEAVAV NAVALAALAAAQVVPVVGEVVDAVLLTEQLVEVLVDVFRA ATAD | WP_079110321.1 L181-D344 |
| 11 | *Streptomyces klenkii* | LPVDWDKVEEKAKATAKKVAQDAEHVEKLPKRSPKGPTWG EAHSTAELSHAKVREVSGAHVAASAAGVGTWVYGMAKTFS DKDATTLDKVAVTGAVVPGLGQALGIADGIQHGDPEAIAV NAVALAALAAAQVVPVVGEVVDAVLLTEQLVEVLVDVFRT ATAD | WP_120757473.1 L227-D390 |
| 12 | *Streptomyces albireticuli* | GINVAAYKADIEKLIKDVYEDPNFSDLKNRTGGPQKDKDT LKGYYERLKPKVETLRPLKAGVSSAVGAAGAISWAIGVAD AFTSENVSSFDKAAAVTAIVPGLGECVGIANAIDKRDPEG LIINTISMAALMASAAVPVLAPIGVALDAGLAAAQGVATV LEYLEIGQ | WP_095582082.1 G232-Q399 |
| 13 | *Streptacidiphilus pinicola* | KVNWKDAYDRSMNIAQDIDGSAEFRATAPARPAAGKAMPE ADVTRLVSTSEDFLEKSTKTSGVKKALKAINNQKMISWGS LLSNALANSHTWSDKNATNLDKAYAVVGGVPVLGEVIGIA SGIDKQDAESIAVNTLSLVGIVAATVCPPLGATVEFVMIG YTAIKLMLSWFTVETI | WP_133259917.1 K246-I421 |
| 14 | *Seinonella peptonophila* | LDWEKIKTESQRIVKQIIEEHPELQSHSKNAVTDKEKLQK IYNDYAPKIDKLSSLKEGVSRATTALNIASWAAGLAETFS NKNADGLDKAAAVTAIIPGLGQAVGIANGIEKHDGEAIAI NSIALSALVVAQAIPIVGEIADVVGAGLILAGGLAQLIQS VSPDT | WP_073156187.1 L252-T416 |

| | | | |
|---|---|---|---|
| 15 | *Longimycelium tulufanense* | ERKTLDKLDRKKIETRAKQVISKLAQDGDLRRELPRRTTT GHAHEEVRGVLAKSRGALHEIHGAATTVAMPLATVAWVED MARVFREKNATTLDKAATVSEIVPVAGQVLGMADGIAHRD AETVAVNAVVLAAIAVSQAVPVVGELVDLGLTAYAVVDVV VRLFGPAREVPITQESYW | WP_189053160.1 E249-W426 |
| 16 | *Austwickia sp. TVS 96-490-7B* (CT3-T) | LSNIKWERVRQKSEELIKKLKDDENVKKAVQERKEGVKPT SDDLQNLHKALIDHESFKELKSVHSNGVKAMDAANAALWG ANVARVFSDSKSDGLEKATAALAAVPGLGQVMGVADGVTH HNTEEVVVQSVALAGFIAAQAIPVVGEIVDIGVLAYQFVE GVIDLSRQMMTSNAR | WP_219106995.1 L233-R407 |
| 17 | *Austwickia chelonae LK16-18* (CT2-T) | LSKVKWEQVREKSKKIIDNVKDNPEFMKKLSAHHERGSAP TTEKITALHNELLDHESFSALKGARSSAGTAATAASAAAW GLAVAQAFTNPKADDLTKATAVLSAVPGLGQALGIADGIK HHNTEEIVVQSISLTALIAAQAIPVVGELVDFGLLAYQLV ESIIDLTRQLSVITAN | WP_116115734.1 L267-N442 |
| 36 | *Klebsiella aerogenes* | GLNWGSIREKAVDKAKSVTSDKKFINSLPERGKNSPANKN DINTIYESSANHSTLKGGSEVIHKASLLLWVANTADVLTS NSSSTLDKATAIIGIVPGIGDAFSLADSIEHDDVEGIITS SIALAAFTVAQAVPVVGELVDLALLTEAAISSLVNVISNY IQSTQR | EIZ2913133.1 G211-R376 |
| 37 | *Streptomyces sp. MUM 178J* | ADLDWEKVRSRAKETARQVAQDSEHVSGLPERRAEGLTKS EAHAVAERTHAKVSEVSGTRVAATAFGVGTWAYGMAQTFT DQNATTLDKVAVTTAIVPGVGNALGIADGIEHHDPEAVAV NAVALAALVAAQAVPVVGELVDTVLLTEQIVETFVNIFQR ATAA | MCH0551590.1 A229-A392 |
| 38 | *Crossiella cryophila* | ETADSLGELFHQGMTIAGEAVAELDLEEALKSQKVAKGDA RAVVAGKVAAVGKQVGKSSALAKFKEVLKPSKGVIALWAV SALYTLAQEDVSTVDKAEAVLAIVPILGPLFSFGNSYRKE DVEGMVISTVALAATLIAFLCPPVGAGIGVGLAVYQVVRA FC | MBB4677777.1 E56-C217 |
| 39 | *Allokutzneria sp. NRRL B-24872* | SLVKGKTDAQAVSAVKQLFENSKTVATSLVADSDMADAIK DAKKAAKVGVAPPVDKVKKVTDPLVEKALKSKSFANTKAG KYLKNAQPSKAGTALWAATTVATFLDGSSSASDKAVAAVG IIPVVGQLASLIHTTIKDPSDVEGIVTGTIGLAAAVITIA C | WP_143261759.1 S60-C220 |
| 40 | *Allokutzneria albata* | KAIGATRESARKFVANILSKSEKATTTLAASNDYKSSMKGF SESFAKLRKAAAETKKPIPTADIKKLLTPHLDKVSTVKDFK KYFTNLKPSQGSAATWLVNLAATVANEESSKADVAVAAAGI IPVVGQVAAIANSVSQGDVEGTVTSAIALAANIAIAVGAG | WP_156051914.1 K78-G240 |
| 41 | *Streptomyces sp. AV19* | CDIGVRDHAEVARKHQERLRAHGDDVEYVTTLDEFEKNPK SLLSPLRETNGVLNDAAAAELGAKDFEAFSTRVTEALGRD LRAVHGLSTDTTLLAKVARQAGRAAELGAKALPYVGIAAT GYAIAEDVKAGDYANLAFDSVAEGLQVAMVAQPELVPLLE PVLFAEQLAQIVYDEIAGWFKRQEQLDHDRKQWDSAAEDL | WP_199893204.1 C260-P481 |

| | | | |
|---|---|---|---|
| | | VEHRDGQWRDHLVGQALQKLFP | |
| 42 | *Streptomyces sp. NRBC_110611* | CGWGDDGVAHTQRERLRAQGGDVEYVTSLEEFEKNPKALL SPLRETNKVLNDAASAELGAKDFETFSARVGAAFSRDLRA VQGLTTDTGLLAGVGRQVGRAGDLGVKVLPYVGIAVTGYA VAEDAKAGDYANLAFDSVAEGLQVAMVAQPEVAPLLEPAL LAEQLAQLVYNEVAGWIAEQKQLEHDRAQWDGAVKDLVAH RDTQWRDRLVGQAVQELFP | WP_147264604.1 C231-P449 |
| 43 | *Streptomyces syringium* | CGIAAADDAKVARTHQERLRAQGADVEYVTSLEEFEKNPK ALMSPLRETSKVLNGAAAAELGAKDFEAFSTQVTEALGRD LKAVHGLTTDTGLLAEVGRRVGRAGELGVKILPYAGIAAT GYAVAEDAKAGDYGNLAFDSVAEGLQVAMVAQPELTPLLE PALLAEQLAQLVYDEVAGWIAEQKQLAHDRGQWDDAVKDL VAHRDTQWRNRLEGQALQELFP | WP_209513619.1 C232-P453 |
| 44 | *Pseudonocardiaceae bacterium YIM PH 21723* | RIGKAALAKATAEAQADAQKILKDPSFKPPHPAPHQKLTD AELTQVANDLAKRVSALHPGARDLTGLGKGLAKALKGVIS GVTAGIQAAEIAESAAGSNSGDLEKVCSALGPIPIVGDIA GIVNAAAEGDAVEIVKNVLTLGLSMVSVAFPPLAPVADAV AFVLEVAVFLGKYLWGLLF | RJQ69589.1 R276-F454 |
| 45 | *Actinokineospora bangkokensis* | SLSQDQLDQAVADAQATAHQVIDDPRSFELPELAKGQVLS EQELAALANGVAQDVAAANSGGERAIVPAYVTRALSKAIP AIAEHGHDAFLKVSAVEGVLTEGNGALAKVTSVVGPVPVL GSLAGIVNEAVGISADDPPAAIAQKALEIAKNVAGVALTV MAIAFPPFAPVAALATAAWEVGAALGKYLLGFLL | WP_143218892.1 S322-L515 |
| 46 | *Streptomyces eurocidicus* | CETADGGDTAEEGKQPGEGRQATESDSTEGRNPQGEEGPA EAGKPVADEEWVARVQRERLARHGSDVEYVTSVAELEKNP KAVLSPLRETNTTLNKARDKPLTAAEYEAFASKAAEAMSK DLKGVQAASTDSTVLAKVGAKLARGGSIAVKVGRPVLKRR LLRSSAASSR | MBF6055834.1 C280-R449 |
| 47 | *Streptomyces pathocidini* | GTGRQDGGFSSEGREGAVTKEDLKGDPQKVVELNRSRIAR HGADVDYVTSLDALAANPKAVLIPLRNSSQTLDAALGKSH LTVDEFDRFSAELTQSFGRDLQAVENLSTSEKLLSSIGKG ARAGGEGLLKVVPFAGVAASSAALAEDLKSGENGDA | WP_169790908.1 G270-A425 |
| 48 | *Streptomyces caatingaensis* | SADIARTHRERLRAHGDDVEYVTSLAEFEKNPKALLSPLR ETGKVLNDAGNAGLSAGDFETFSARVGEAFERDLRAAHEL TTDTGLLARVGAEAGRAAELGAKALPFVGIAATGYAVAED VKSGDYANLAFDSVAEGLQVAMTAQPEFA | WP_157868472.1 S241-A389 |
| **Constructs and Modules** | | | |
| **SEQ ID NO:** | **Description** | **Sequence** | **Source** |
| 18 | *Vibrio vulnificus* (RRSP) | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG | WP_011081430.1 G3580–L4089 |

| | | | |
|---|---|---|---|
| | | FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSL | |
| 19 | Her3 affibody<br>(ZHer3:08699) | AEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQ<br>SSELLSEAKKLNDSQAPK | Nazari et al.<br>(2019) |
| 20 | A20 peptide<br>(A20FMDV2) | NAVPNLRGDLQVLAQKVART | DiCara et al.<br>(2007) |
| 22 | RRSP – DT-T -<br>ZHer3-A20<br>Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS<br>CAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGP<br>IKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKT<br>VTGTNPVFAGANYAAWAVNVAQVIDSETADNLEKTTAALS<br>ILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIP<br>LVGELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHK<br>TQPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQLPNLTKYQ<br>KAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGG<br>GSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK | n/a |
| 23 | RRSP – CT1-T -<br>ZHer3-A20<br>Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV | n/a |

| | | | |
|---|---|---|---|
| | | ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS<br>CAGNRVRRSVGSSLSCLSKINWKNVREKADALTKKVHADK<br>EFMDKLSTHHQRGEAPSVEKTTALHNALLEHESFSALKGA<br>RASGKVGAAASTAAWGVAVAQAFTDPKADALTKTAATLSV<br>VPGLGQALGIADGIKHENTEEIVVQSISLAGLLAAQAIPV<br>VGEAVDFGLLVYQLVETIVDLATHLSSAAANPAYSPGHKT<br>QPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQLPNLTKYQK<br>AAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGGG<br>SNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK | |
| 24 | RRSP – CT2-T - ZHer3-A20 Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS<br>CAGNRVRRSVGSSLSCLSKVKWEQVREKSKKIIDNVKDNP<br>EFMKKLSAHHERGSAPTTEKITALHNELLDHESFSALKGA<br>RSSAGTAATAASAAAWGLAVAQAFTNPKADDLTKATAVLS<br>AVPGLGQALGIADGIKHHNTEEIVVQSISLTALIAAQAIP<br>VVGELVDFGLLAYQLVESIIDLTRQLSVITANPAYSPGHK<br>TQPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQLPNLTKYQ<br>KAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGG<br>GSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK | n/a |
| 25 | RRSP – CT3-T - ZHer3-A20 Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN | n/a |

| | | | |
|---|---|---|---|
| | | QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS<br>CAGNRVRRSVGSSLSCLSNIKWERVRQKSEELIKKLKDDE<br>NVKKAVQERKEGVKPTSDDLQNLHKALIDHESFKELKSVH<br>SNGVKAMDAANAALWGANVARVFSDSKSDGLEKATAALAA<br>VPGLGQVMGVADGVTHHNTEEVVVQSVALAGFIAAQAIPV<br>VGEIVDIGVLAYQFVEGVIDLSRQMMTSNARPAYSPGHKT<br>QPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQLPNLTKYQK<br>AAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGGG<br>SNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK | |
| 26 | RRSP – T (S. klenkii) - ZHer3-A20 Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILD<br>QLDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQA<br>LTQLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIF<br>MVAIEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETP<br>SDLGFDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLD<br>ESTKTVHLGYSYQELQDLTGAESVQMAFYFLKEAAKKAD<br>PISGDSAEMILLKKFADQSYLSQLDSDRMDQIEGIYRSS<br>HETDIDAWDRRYSGTGYDELTNKLASATGVDEQLAVLLD<br>DRKGLLIGEVHGSDVNGLRFVNEQMDALKKQGVTVIGLE<br>HLRSDLAQPLIDRYLATGVMSSELSAMLKTKHLDVTLFE<br>NARANGMRIVALDANSSARPNVQGTEHGLMYRAGAANNI<br>AVEVLQNLPDGEKFVAIYGKAHLQSHKGIEGFVPGITHR<br>LDLPALKVSDSNQFTVEQDDVSLRVVYDDVANKPKITFK<br>GSLGGGGSGGGGSCAGNRVRRSVGSSLSCLPVDWDKVEE<br>KAKATAKKVAQDAEHVEKLPKRSPKGPTWGEAHSTAELS<br>HAKVREVSGAHVAASAAGVGTWVYGMAKTFSDKDATTLD<br>KVAVTGAVVPGLGQALGIADGIQHGDPEAIAVNAVALAA<br>LAAAQVVPVVGEVVDAVLLTEQLVEVLVDVFRTATADPA<br>YSPGHKTQPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQL<br>PNLTKYQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPK<br>GGGGSGGGGSNAVPNLRGDLQVLAQKVARTRQALVPRGS<br>AWSHPQFEK | n/a |
| 27 | RRSP – T (S. sp TLI053) - ZHer3-A20 Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS | n/a |

| | | | |
|---|---|---|---|
| | | CAGNRVRRSVGSSLSCIPAASWEAIEHRSKEMASAVARDT<br>EYTKSLPNRHPKGPTWSEAHTTSTSTHARVSAKSGAHVAV<br>GAFAVGSWIYGMSETFANKNVTTLDKAAATVAIVPGIGHA<br>LGIAAALEHHDIEGVVVNAISIAALAAAQVVPVVGEIVDA<br>ALLAEQLVEVLVHVFRASTTEPAYSPGHKTQPFGGGGSGG<br>GGSAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDD<br>PSQSSELLSEAKKLNDSQAPKGGGGSGGGGSNAVPNLRGD<br>LQVLAQKVARTRQALVPRGSAWSHPQFEK | |
| 28 | RRSP – T (L. tulufanense) - ZHer3-A20 Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS<br>CAGNRVRRSVGSSLSCERKTLDKLDRKKIETRAKQVISKL<br>AQDGDLRRELPRRTTTGHAHEEVRGVLAKSRGALHEIHGA<br>ATTVAMPLATVAWVEDMARVFREKNATTLDKAATVSEIVP<br>VAGQVLGMADGIAHRDAETVAVNAVVLAAIAVSQAVPVVG<br>ELVDLGLTAYAVVDVVVRLFGPAREVPITQESYWPAYSPG<br>HKTQPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQLPNLTK<br>YQKAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSG<br>GGGSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQF<br>EK | n/a |
| 29 | RRSP – T (S. piniterrae) - ZHer3-A20 Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS<br>CAGNRVRRSVGSSLSCELAKDKARDILSEAGNEVSLPQRD | n/a |

| | | | |
|---|---|---|---|
| | | SDGLSKQEITATAEATRSKLGTGVHGAVSAAMVADWAHDV<br>ARTFADPKATKLDKAAAVTAIAPVIGQAVNIADGIQHHDK<br>KTIVVNSLVLAAVVAAQAVPAVGEVVDAAIVADFVVEKLV<br>GWFTPTAKPGPEHVAPAYSPGHKTQPFGGGGSGGGGSAEA<br>KYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQSSE<br>LLSEAKKLNDSQAPKGGGGSGGGGSNAVPNLRGDLQVLAQ<br>KVARTRQALVPRGSAWSHPQFEK | |
| 30 | RRSP – T (S. pinicola) - ZHer3-A20 Construct | GDKTKVVVDLAQIFTVQELKERAKVFAKPIGASYQGILDQ<br>LDLVHQAKGRDQIAASFELNKKINDYIAEHPTSGRNQALT<br>QLKEQVTSALFIGKMQVAQAGIDAIAQTRPELAARIFMVA<br>IEEANGKHVGLTDMMVRWANEDPYLAPKHGYKGETPSDLG<br>FDAKYHVDLGEHYADFKQWLETSQSNGLLSKATLDESTKT<br>VHLGYSYQELQDLTGAESVQMAFYFLKEAAKKADPISGDS<br>AEMILLKKFADQSYLSQLDSDRMDQIEGIYRSSHETDIDA<br>WDRRYSGTGYDELTNKLASATGVDEQLAVLLDDRKGLLIG<br>EVHGSDVNGLRFVNEQMDALKKQGVTVIGLEHLRSDLAQP<br>LIDRYLATGVMSSELSAMLKTKHLDVTLFENARANGMRIV<br>ALDANSSARPNVQGTEHGLMYRAGAANNIAVEVLQNLPDG<br>EKFVAIYGKAHLQSHKGIEGFVPGITHRLDLPALKVSDSN<br>QFTVEQDDVSLRVVYDDVANKPKITFKGSLGGGGSGGGGS<br>CAGNRVRRSVGSSLSCKVNWKDAYDRSMNIAQDIDGSAEF<br>RATAPARPAAGKAMPEADVTRLVSTSEDFLEKSTKTSGVK<br>KALKAINNQKMISWGSLLSNALANSHTWSDKNATNLDKAY<br>AVVGGVPVLGEVIGIASGIDKQDAESIAVNTLSLVGIVAA<br>TVCPPLGATVEFVMIGYTAIKLMLSWFTVETIPAYSPGHK<br>TQPFGGGGSGGGGSAEAKYAKEKYNAYYEIWQLPNLTKYQ<br>KAAFIGKLQDDPSQSSELLSEAKKLNDSQAPKGGGGSGGG<br>GSNAVPNLRGDLQVLAQKVARTRQALVPRGSAWSHPQFEK | n/a |
| 31 | $DT_{1-389}$-ZHer3-A20 Construct | GADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQKPKS<br>GTQGNYDDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGV<br>VKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGT<br>EEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALS<br>VELEINFETRGKRGQDAMYEYMAQACAGNRVRRSVGSSLS<br>CINLDWDVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSE<br>EKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAA<br>WAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGA<br>VHHNTEEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNF<br>VESIINLFQVVHNSYNRPAYSPGHKTQPFGGGGSGGGGSA<br>EAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPSQS<br>SELLSEAKKLNDSQAPKGGGGSGGGGSNAVPNLRGDLQVL<br>AQKVARTRQALVPRGSAWSHPQFEK | n/a |
| 32 | $CT2_{1-391}$-ZHer3-A20 Construct | YANDAVIADQSKTVDSFTSYHGAKPESFESVLAGIKKPES<br>GSQGNHDPEWKGFYTTDNKHAAAGYTVSDESVMTGKAGGV<br>VKVTYPGKTRVLAVKPLSATELKTYLGLAADKPLIDQLNN | n/a |

| | | | |
|---|---|---|---|
| | | KDFINKFGEGASRVVLQMPFADGTSDVEYIHNWEDATQLQ VATEVRFDNLGKRGQDEMNRYMNLANCPSVSAVRVKRNPA KLCLSKVKWEQVREKSKKIIDNVKDNPEFMKKLSAHHERG SAPTTEKITALHNELLDHESFSALKGARSSAGTAATAASA AAWGLAVAQAFTNPKADDLTKATAVLSAVPGLGQALGIAD GIKHHNTEEIVVQSISLTALIAAQAIPVVGELVDFGLLAY QLVESIIDLTRQLSVITANPPTEVTHSSKLAGGGGSGGGG SAEAKYAKEKYNAYYEIWQLPNLTKYQKAAFIGKLQDDPS QSSELLSEAKKLNDSQAPKGGGGSGGGGSNAVPNLRGDLQ VLAQKVARTRQALVPRGSAWSHPQFEK | |
| 33 | Linker from *C. diphtheriae* containing furin recognition site (with bracketing cysteines) | CAGNRVRRSVGSSLSC | n/a |
| 34 | Linker after T domain, from *C. diphtheriae* | PAYSPGHKTQPF | n/a |

**[00550]** **References**

**[00551]** Orrell, K. E., Mansfield, M. J., Doxey, A. C. & Melnyk, R. A. The C. difficile toxin B membrane translocation machinery is an evolutionarily conserved protein delivery apparatus. *Nature Communications* **11**, 1–11 (2020).

**[00552]** Park, M. *et al.* Intracellular Delivery of Human Purine Nucleoside Phosphorylase by Engineered Diphtheria Toxin Rescues Function in Target Cells. *Molecular Pharmaceutics* **15**, 5217–5226 (2018).

**[00553]** Nazari, M.; Zamani Koukhaloo, S.; Mousavi, S.; Minai-Tehrani, A.; Emamzadeh, R.; Cheraghi, R. Development of a ZHER3-Affibody-Targeted Nano-Vector for Gene Delivery to HER3-Overexpressed Breast Cancer Cells. Macromol. Biosci. 2019, 19 (11).

**[00554]** DiCara, D.; Rapisarda, C.; Sutcliffe, J. L.; Violette, S. M.; Weinreb, P. H.; Hart, I. R.; Howard, M. J.; Marshall, J. F. Structure-Function Analysis of Arg-Gly-Asp Helix Motifs in Avβ6 Integrin Ligands. J. Biol. Chem. 2007, 282 (13), 9657–9665.

**[00555]** Sugiman-Marangos, S.N., Gill, S.K., Mansfield, M.J. et al. Structures of distant diphtheria toxin homologs reveal functional determinants of an evolutionarily conserved toxin scaffold. Commun Biol 5, 375 (2022).

**[00556]** In the preceding description, for purposes of explanation, numerous details are set forth in order to provide a thorough understanding of the embodiments. However, it will be apparent to one skilled in the art that these specific details are not required. In other

instances, well-known electrical structures and circuits are shown in block diagram form in order not to obscure the understanding. For example, specific details are not provided as to whether the embodiments described herein are implemented as a software routine, hardware circuit, firmware, or a combination thereof.

**[00557]** The above-described embodiments are intended to be examples only. Alterations, modifications and variations can be effected to the particular embodiments by those of skill in the art. The scope of the claims should not be limited by the particular embodiments set forth herein, but should be construed in a manner consistent with the specification as a whole.

**CLAIMS:**

1. A recombinant polypeptide of general formula (I):

**A-B-C** (i)

wherein:

**A** is a cargo molecule;

**B** is a translocation polypeptide comprising:

a) a translocation domain from:

i. the *Austwickia chelonae* protein of SEQ ID NO: 2,

ii. the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1,

iii. the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1,

iv. the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1,

v. the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1,

vi. the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1,

vii. the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1,

viii. the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1,

ix. the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1,

x. the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1,

xi. the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1,

xii. the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1,

xiii. the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1,

xiv. the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1,

xv. the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1;

xvi. the *Klebsiella aerogenes* protein of GenBank Accession EIZ2913133.1,

xvii. the *Streptomyces sp. MUM 178J* protein of GenBank Accession MCH0551590.1,

xviii. the *Crossiella cryophila* protein of GenBank Accession MBB4677777.1,

xix. the *Allokutzneria sp. NRRL B-24872* protein of GenBank Accession WP_143261759.1,

xx. the *Allokutzneria albata* protein of GenBank Accession WP_156051914.1,

xxi. the *Streptomyces sp. AV19* protein of GenBank Accession WP_199893204.1,

xxii. the *Streptomyces sp. NRBC_110611* protein of GenBank Accession WP_147264604.1,

xxiii. the *Streptomyces syringium* protein of GenBank Accession WP_209513619.1,

xxiv. the *Pseudonocardiaceae bacterium YIM PH 21723* protein of GenBank Accession RJQ69589.1,

xxv. the *Actinokineospora bangkokensis* protein of GenBank Accession WP_143218892.1,

xxvi. the *Streptomyces eurocidicus* protein of GenBank Accession MBF6055834.1,

xxvii. the *Streptomyces pathocidini* protein of GenBank Accession WP_169790908.1, or

xxviii. the *Streptomyces caatingaensis* protein of GenBank Accession WP_157868472.1,or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a); and

**C** is a targeting moiety.

2. A recombinant polypeptide of general formula (I):

**A-B-C** (i)

wherein:

**A** is a cargo molecule;

**B** is a translocation polypeptide comprising:

a) a translocation domain from:

i. the *Austwickia chelonae* protein of SEQ ID NO: 2,

ii. the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1,

iii. the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1,

iv. the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1,

v. the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1,

vi. the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1,

vii. the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1,

viii. the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1,

ix. the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1,

x. the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1,

xi. the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1,

xii. the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1, or

xiii. the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1

b) a translocation domain that is at least 80% identical to the translocation domain defined in a);

c) a translocation domain comprising an amino acid sequence according to any one of SEQ ID Nos: 36 to 48; or

d) a translocation domain that is at least 80% identical to the translocation domain defined in c), and

**C** is a targeting moiety.

3. A recombinant polypeptide of general formula (I):

**A-B-C** (i)

wherein:

**A** is a cargo molecule;

**B** is a translocation polypeptide comprising:

a) a translocation domain from:

i. the *Austwickia chelonae* protein of SEQ ID NO: 2,

ii. the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1,

iii. the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1,

iv. the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1, or

v. the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1; or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a); and

**C** is a targeting moiety.

4. A recombinant polypeptide of general formula (I):

**A-B-C** (i)

wherein:

**A** is a cargo molecule;

**B** is a translocation polypeptide comprising:

a) a translocation domain from:

i. the *Austwickia chelonae* protein of SEQ ID NO: 2,

ii. the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1; or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a); and

**C** is a targeting moiety.

5. The recombinant polypeptide of any one of claims 1 to 4,

wherein in a):

the translocation domain from the *Austwickia chelonae* protein has the amino acid sequence of SEQ ID NO: 3,

the translocation domain from the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1 has the amino acid sequence of SEQ ID NO: 4,

the translocation domain from the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1 has the amino acid sequence of SEQ ID NO: 5,

the translocation domain from the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1 has the amino acid sequence of SEQ ID NO: 6,

the translocation domain from the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1 has the amino acid sequence of SEQ ID NO: 7,

the translocation domain from the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1 has the amino acid sequence of SEQ ID NO: 8,

the translocation domain from the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1 has the amino acid sequence of SEQ ID NO: 9,

the translocation domain from the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1 has the amino acid sequence of SEQ ID NO: 10,

the translocation domain from the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1 has the amino acid sequence of SEQ ID NO: 11,

the translocation domain from the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1 has the amino acid sequence of SEQ ID NO: 12,

the translocation domain from the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1 has the amino acid sequence of SEQ ID NO: 13,

the translocation domain from the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1 has the amino acid sequence of SEQ ID NO: 14,

the translocation domain from the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1 has the amino acid sequence of SEQ ID NO: 15,

the translocation domain from the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1 has the amino acid sequence of SEQ ID NO: 16, and

the translocation domain from the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_162873017.1 has the amino acid sequence of SEQ ID NO: 17.

6. The recombinant polypeptide of any one of claims 1 to 4, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Austwickia chelonae* protein, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 3, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

7. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptosporangium nondiastaticum* protein of GenBank Accession PSJ28985.1 wherein the translocation domain has the amino acid sequence of SEQ ID NO: 4, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

8. The recombinant polypeptide of any one of claims 1 to 3, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces sp.TLI 053* protein of GenBank Accession SDT83331.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 5, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

9. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces sp. SLBN-118* protein of GenBank Accession WP_160159328.1 wherein the translocation domain has the amino acid sequence of SEQ ID NO: 6, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

10. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces sp. AA8* protein of GenBank Accession WP_168096531.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 7, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

11. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces roseoverticillatus* protein of GenBank Accession WP_078659863.1 wherein the translocation domain has the amino acid sequence of SEQ ID NO: 8, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

12. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces piniterrae* protein of GenBank Accession JZ58907.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 9, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

13. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces MBT76* protein of GenBank Accession WP_079110321.1 wherein the translocation domain has the amino acid sequence of SEQ ID NO: 10, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

14. The recombinant polypeptide of any one of claims 1 to 3, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces klenkii* protein of GenBank Accession WP_120757473.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 11, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

15. The recombinant polypeptide of claim 1, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptomyces albireticuli* protein of GenBank Accession WP_095582082.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 12, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

16. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Streptacidiphilus pinicola* protein of GenBank Accession WP_133259917.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 13, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

17. The recombinant polypeptide of claim 1, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Seinonella peptonophila* protein of GenBank Accession WP_073156187.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 14, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

18. The recombinant polypeptide of claim 1 or 2, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Longimycelium tulufanense* protein of GenBank Accession WP_189053160.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 15, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

19. The recombinant polypeptide of any one of claims 1 to 3, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Austwickia sp. TVS 96-490-7B* protein of GenBank Accession WP_219106995.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 16, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

20. The recombinant polypeptide of any one of claims 1 to 4, wherein the translocation polypeptide comprises:

a) the translocation domain from the *Austwickia chelonae LK16-18* protein of GenBank Accession WP_116115734.1, wherein the translocation domain has the amino acid sequence of SEQ ID NO: 17, or

b) a translocation domain that is at least 80% identical to the translocation domain defined in a).

21. The recombinant polypeptide of any one of claims 1 to 20, wherein the translocation domain is as defined in b) and is at least at least 85% identical to the translocation domain defined in a) across the full length thereof.

22. The recombinant polypeptide of any one of claims 1 to 20, wherein the translocation domain is as defined in b) and is at least at least 90% identical to the translocation domain defined in a) across the full length thereof.

23. The recombinant polypeptide of any one of claims 1 to 20, wherein the translocation domain is as defined in b) and is at least at least 95% identical to the translocation domain defined in a) across the full length thereof.

24. The recombinant polypeptide of any one of claims 1 to 20, wherein the translocation domain is as defined in b) and is at least at least 98% identical to the translocation domain defined in a) across the full length thereof.

25. The recombinant polypeptide of any one of claims 1 to 20, wherein the translocation domain is as defined in b) and is at least at least 99% identical to the translocation domain defined in a) across the full length thereof.

26. The recombinant polypeptide of any one of claims 1 to 20, wherein the translocation polypeptide is as defined in a).

27. The recombinant polypeptide of any one of claims 1 to 26, wherein A and B are separated by an amino acid linker.

28. The recombinant polypeptide of claim 27 or 28, wherein the amino acid linker comprises $(G_4S)_2$.

29. The recombinant polypeptide of claim 27 or 28, wherein the amino acid linker is cleavable, preferably wherein the amino acid linker comprises a protease recognition site, preferably wherein the protease recognition site is bracketed by cysteine residues, more preferably wherein the protease recognition site is a furin protease recognition site

30. The recombinant polypeptide of claim 29, wherein the amino acid linker comprises SEQ ID NO: 32, which comprises the furin protease recognition site.

31. The recombinant polypeptide of claim 30, wherein said $(G_4S)_2$ is positioned N-terminally with respect to said SEQ ID NO: 32.

32. The recombinant polypeptide of any one of claims 1 to 27, wherein B and C are separated by an amino acid linker.

33. The recombinant polypeptide of claim 27 or 28, wherein the amino acid linker comprises $(G_4S)_2$.

34. The recombinant polypeptide of claim 32 or 33, wherein the amino acid linker comprises SEQ ID NO: 33.

35. The recombinant polypeptide of claim 34, wherein said $(G_4S)_2$ is positioned N-terminally with respect to said SEQ ID NO: 33.

36. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 821 of SEQ ID NO: 22, preferably 90% identical to amino acids 1 to 821 of SEQ ID NO: 22, more preferably 95% identical to amino acids 1 to 821 of SEQ ID NO: 22, even more preferably 100% identical to amino acids 1 to 821 of SEQ ID NO: 22.

37. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 22.

38. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 822 of SEQ ID NO: 23, preferably 90% identical to amino acids 1 to 822 of SEQ ID NO: 23, more preferably 95% identical to amino acids 1 to 822 of SEQ ID NO: 23, even more preferably 100% identical to amino acids 1 to 822 of SEQ ID NO: 23.

39. The recombinant polypeptide of any one of claims 1 to 4, which the sequence of SEQ ID NO: 23.

40. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 822 SEQ ID NO: 24, preferably 90% identical to amino acids 1 to 822 SEQ ID NO: 24, more preferably 95% identical to amino acids 1 to 822 SEQ ID NO: 24, even more preferably 100% identical to amino acids 1 to 822 SEQ ID NO: 24.

41. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 24.

42. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 821 of SEQ ID NO: 25, preferably 90% identical to amino acids 1 to 821 of SEQ ID NO: 25, more preferably 95% identical to amino acids 1 to 821 of SEQ ID NO: 25, even more preferably 100% identical to amino acids 1 to 821 of SEQ ID NO: 25.

43. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 25.

44. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 806 of SEQ ID NO: 26, preferably 90% identical to amino acids 1 to 806 of SEQ ID NO: 26, more preferably 95% identical to amino acids 1 to 806 of SEQ ID NO: 26, even more preferably 100% identical to amino acids 1 to 806 of SEQ ID NO: 26.

45. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 26.

46. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 811 of SEQ ID NO: 27, preferably 90% identical to amino acids 1 to 811 of SEQ ID NO: 27, more preferably 95% identical to amino acids 1 to 811 of SEQ ID NO: 27, even more preferably 100% identical to amino acids 1 to 811 of SEQ ID NO: 27.

47. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 27.

48. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 822 of SEQ ID NO: 28, preferably 90% identical to amino acids 1 to 822 of SEQ ID NO: 28, more preferably 95% identical to amino acids 1 to 822 of SEQ ID NO: 28, even more preferably 100% identical to amino acids 1 to 822 of SEQ ID NO: 28.

49. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 28.

50. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 796 of SEQ ID NO: 29, preferably 90% identical to amino acids 1 to 796 of SEQ ID NO: 29, more preferably 95% identical to amino acids 1 to 796 of SEQ ID NO: 29, even more preferably 100% identical to amino acids 1 to 796 of SEQ ID NO: 29.

51. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 29.

52. The recombinant polypeptide of any one of claims 1 to 4, comprising an amino acid sequence that is at least 80% identical to amino acids 1 to 822 of SEQ ID NO: 30, preferably 90% identical to amino acids 1 to 822 of SEQ ID NO: 30, more preferably 95% identical to amino acids 1 to 822 of SEQ ID NO: 30, even more preferably 100% identical to amino acids 1 to 822 of SEQ ID NO: 30.

53. The recombinant polypeptide of any one of claims 1 to 4, which comprises the sequence of SEQ ID NO: 30.

54. The recombinant polypeptide of any one of claims 1 to 35, wherein the targeting moiety comprises a targeting polypeptide or aptamer.

55. The recombinant polypeptide of claim 48, wherein the targeting polypeptide comprises an antibody, a binding fragment of an antibody, an affibody, a peptide, an affitin, a DARPin, or a receptor ligand.

56. The recombinant polypeptide of claim 55, wherein the targeting polypeptide comprises an affibody against Her3.

57. The recombinant polypeptide of claim 56, wherein the affibody against Her3 comprises the amino acid sequence of SEQ ID NO: 19.

58. The recombinant polypeptide of claim 59, wherein the targeting polypeptide comprises a receptor ligand for $\alpha v\beta 6$ integrin.

59. The recombinant polypeptide of claim 52, wherein the receptor ligand for $\alpha v\beta 6$ integrin comprises the amino acid sequence of SEQ ID NO: 20.

60. The recombinant polypeptide of claim 54, wherein the targeting moiety comprises at least two targeting polypeptides, at least two aptamers, or a combination of a targeting polypeptide and an aptamer.

61. The recombinant polypeptide of claim 60, wherein the at least two targeting polypeptides are selected from the group an antibody, a binding fragment of an antibody, an affibody, a peptide, an affitin, a DARPin, a receptor ligand, and combinations thereof.

62. The recombinant polypeptide of claim 60 or 61, wherein the at least two targeting polypeptides comprise an affibody against Her3.

63. The recombinant polypeptide of claim 62, wherein the affibody against Her3 comprises the amino acid sequence of SEQ ID NO: 19.

64. The recombinant polypeptide of any one of claims 60 to 63, wherein the at least two targeting polypeptides comprise a receptor ligand for $\alpha v\beta 6$ integrin.

65. The recombinant polypeptide of claim 64, wherein the receptor ligand for $\alpha v\beta 6$ integrin comprises the amino acid sequence of SEQ ID NO: 20.

66. The recombinant polypeptide of any one of claims 60 to 65, wherein the at least two least two targeting polypeptides, the at least two aptamers, or the combination are separated by an amino acid linker.

67. The recombinant polypeptide of claim 66, wherein the amino acid linker comprises $(G_4S)_2$.

68. The recombinant polypeptide of any one of claims 1 to 35 and 54 to 67, wherein the targeting moiety binds to a cell surface protein.

69. The recombinant polypeptide of claim 68, wherein the cell surface protein is lineage-specific or tissue-specific.

70. The recombinant polypeptide of claim 67, wherein the cell surface protein is ubiquitously expressed.

71. The recombinant polypeptide of claim 67, wherein the cell surface protein is expressed in a disease cell.

72. The recombinant polypeptide of claim 67, wherein the cell surface protein is specific to a disease cell and is not expressed in a corresponding healthy cell.

73. The recombinant polypeptide of claim 67, wherein the cell surface protein has elevated expression in a disease cell compared to a corresponding healthy cell.

74. The recombinant polypeptide of any one of claims 71 to 73, wherein the disease cell is a cancer cell.

75. The recombinant polypeptide of any one of claims 1 to 35 and 54 to 74, wherein the cargo molecule comprises a therapeutic polypeptide.

76. The recombinant polypeptide of claim 75, wherein the therapeutic polypeptide comprises a cytotoxic polypeptide, preferably a polypeptide toxin or a functional fragment thereof.

77. The recombinant polypeptide of claim 76, wherein the cytotoxic polypeptide comprises a catalytic domain from Diphtheria Toxin.

78. The recombinant polypeptide of claim 76, wherein the cytotoxic polypeptide comprises a catalytic domain from a Chelona Toxin.

79. The recombinant polypeptide of claim 78, wherein the catalytic domain is from:
the Chelona Toxin set forth in SEQ ID NO: 2 (CT1),
the Chelona Toxin set forth in SEQ ID NO: 21 (CT2), or
the Chelona Toxin set forth in SEQ ID NO: 35 (CT3).

80. The recombinant polypeptide of claim 79,
wherein the catalytic domain from the Chelona Toxin set forth in SEQ ID NO: 2 (CT1) has the amino acid sequence according to amino acid positions 1 to 186 of SEQ ID NO: 2;
wherein the catalytic domain from the Chelona Toxin set forth in SEQ ID NO: 21 (CT2) has the amino acid sequence according to amino acid positions 1 to 186 of SEQ ID NO: 21; or
wherein the catalytic domain from the Chelona Toxin set forth in SEQ ID NO: 35 (CT3) has the amino acid sequence according to amino acid positions 1 to 191 of SEQ ID NO: 35.

81. The recombinant polypeptide of claim 75, wherein the therapeutic polypeptide comprises a protein that is deficient is a disease state, or a functional fragment thereof.

82. The recombinant polypeptide of claim 75, wherein the therapeutic polypeptide comprises Ras/Rap1-specific endopeptidase (RRSP) from Vibrio vulnificus (SEQ ID NO: 18).

83. The recombinant polypeptide of any one of claims 1 to 35 and 54 to 74, wherein the cargo molecule comprises an N-terminal cysteine residue for use in “click” chemistry.

84. The recombinant polypeptide of any one of claims 1 to 35, 54 to 74, and 83 wherein the cargo molecule comprises a nucleic acid molecule.

85. A nucleic acid encoding the recombinant polypeptide as defined in any one of claims 1 to 81.

86. A composition comprising the recombinant polypeptide as defined in any one of claims 1 to 82, together with an acceptable excipient, diluent, or carrier.

87. A pharmaceutical composition comprising the recombinant polypeptide as defined in any one of claims 1 to 82, together with a pharmaceutically acceptable excipient, diluent, or carrier.

88. A method of delivering a cargo molecule to a cell comprising contacting the cell with the recombinant polypeptide as defined in any one of claims 1 to 82.

89. A use of the recombinant polypeptide as defined in any one of claims 1 to 82 for delivery of the cargo molecule to a cell.

90. The recombinant polypeptide as defined in any one of claims 1 to 82 for use in delivery of the cargo molecule to a cell.

91. A method treating cancer in a subject comprising administering to the subject the recombinant polypeptide as defined in any one of claims 1 to 82.

92. A use of the recombinant polypeptide as defined in any one of claims 1 to 82 for treatment of cancer in a subject.

93. The recombinant polypeptide as defined in any one of claims 1 to 82 for use in treatment of cancer in a subject.

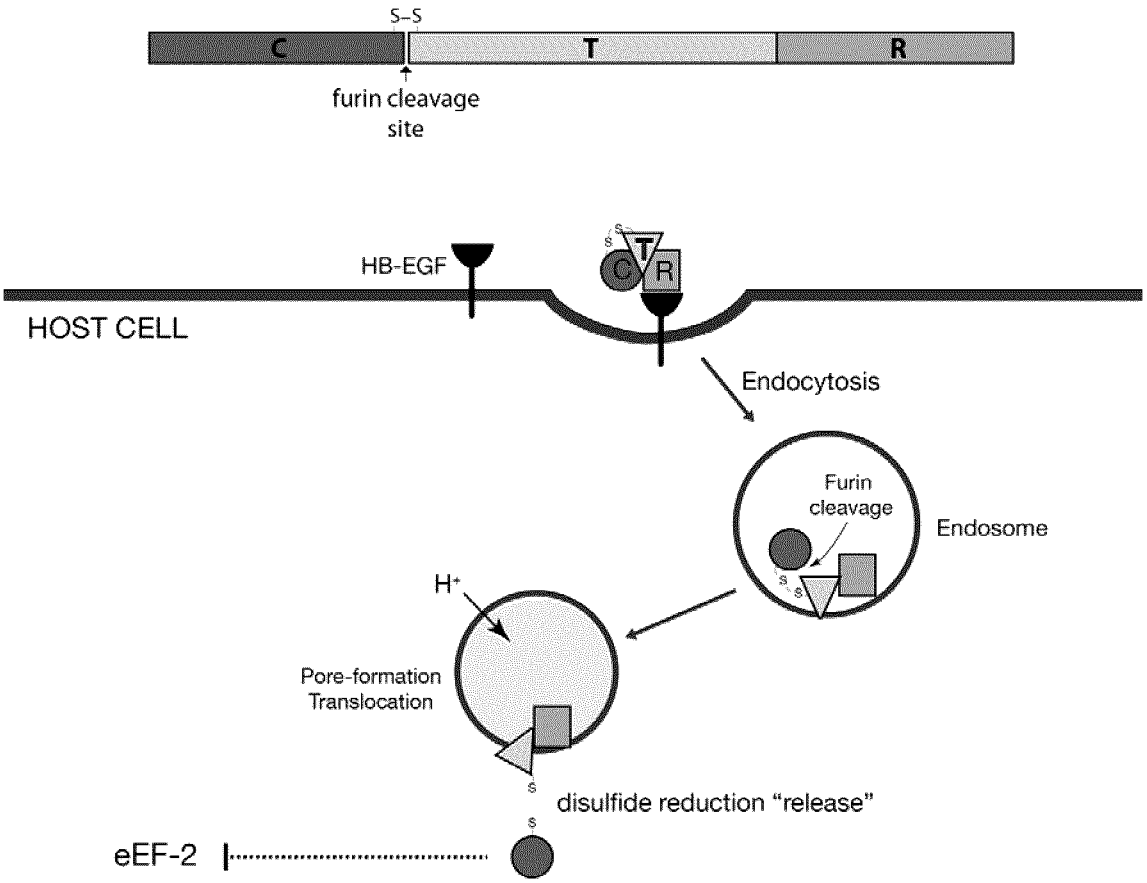

S–S
C
T
R
furin cleavage site
HB-EGF
HOST CELL
Endocytosis
Furin cleavage
Endosome
H+
Pore-formation Translocation
disulfide reduction "release"
eEF-2

**Figure 1A**

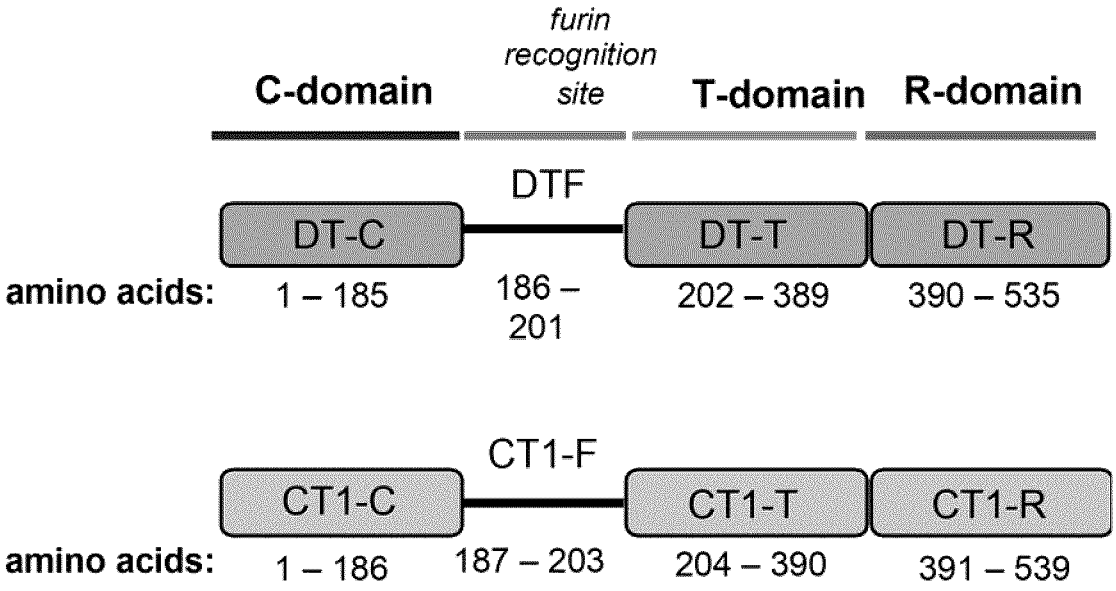

furin recognition site
C-domain
T-domain
R-domain
DTF
DT-C
DT-T
DT-R
amino acids:
1 – 185
186 – 201
202 – 389
390 – 535
CT1-F
CT1-C
CT1-T
CT1-R
amino acids:
1 – 186
187 – 203
204 – 390
391 – 539

**Figure 1B**



**Diphtheria Toxin (DT)**
**(PDB ID: 1MDT)**



**Chelona Toxin 1 (CT1)**
DT-like (*Austwickia chelonae*)
Resolution: 2.50Å
RMSD to DT: 2.4Å

**Figure 1C**

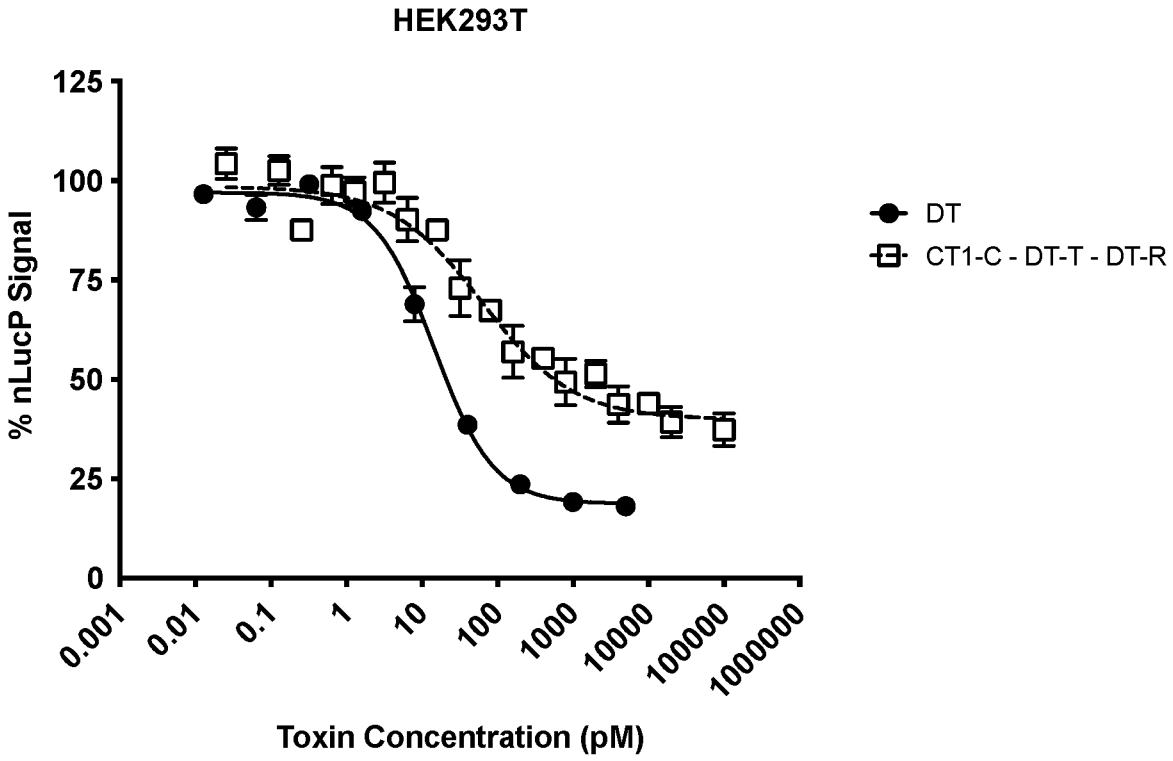

HEK293T
% nLucP Signal
125
100
75
50
25
0
0.001
0.01
0.1
1
10
100
1000
10000
100000
1000000
Toxin Concentration (pM)
DT
CT1-C - DT-T - DT-R

Figure 2A

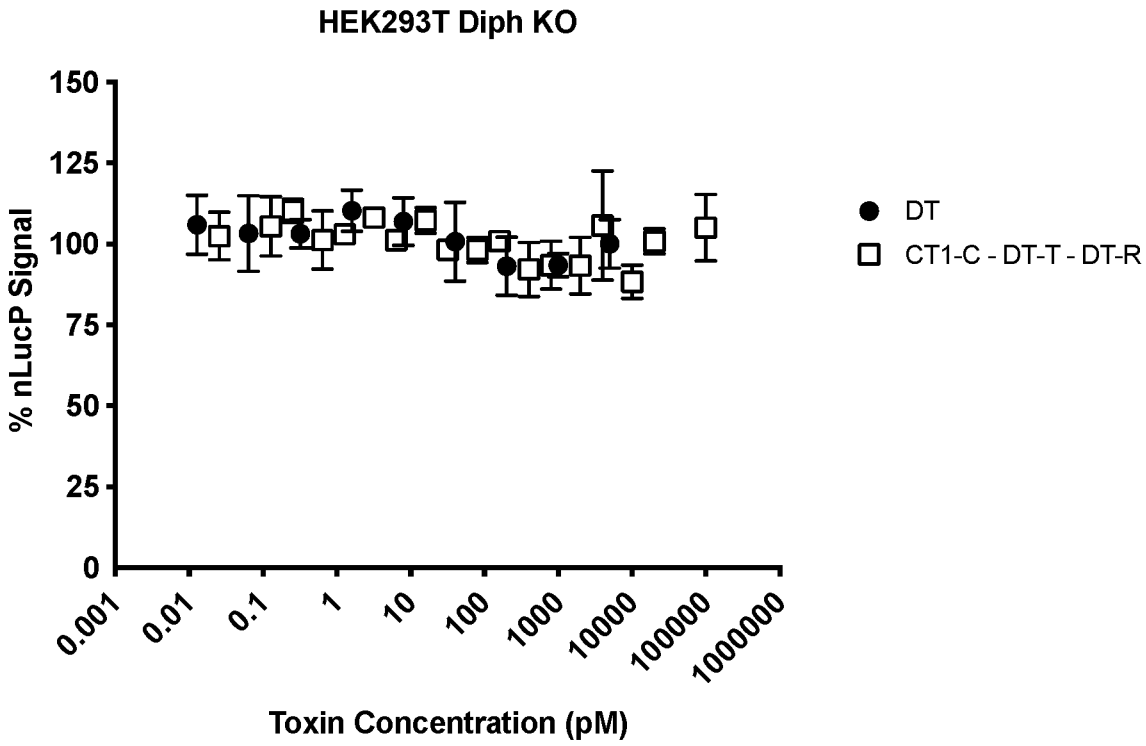

HEK293T Diph KO
% nLucP Signal
150
125
100
75
50
25
0
0.001
0.01
0.1
1
10
100
1000
10000
100000
1000000
Toxin Concentration (pM)
DT
CT1-C - DT-T - DT-R

Figure 2B

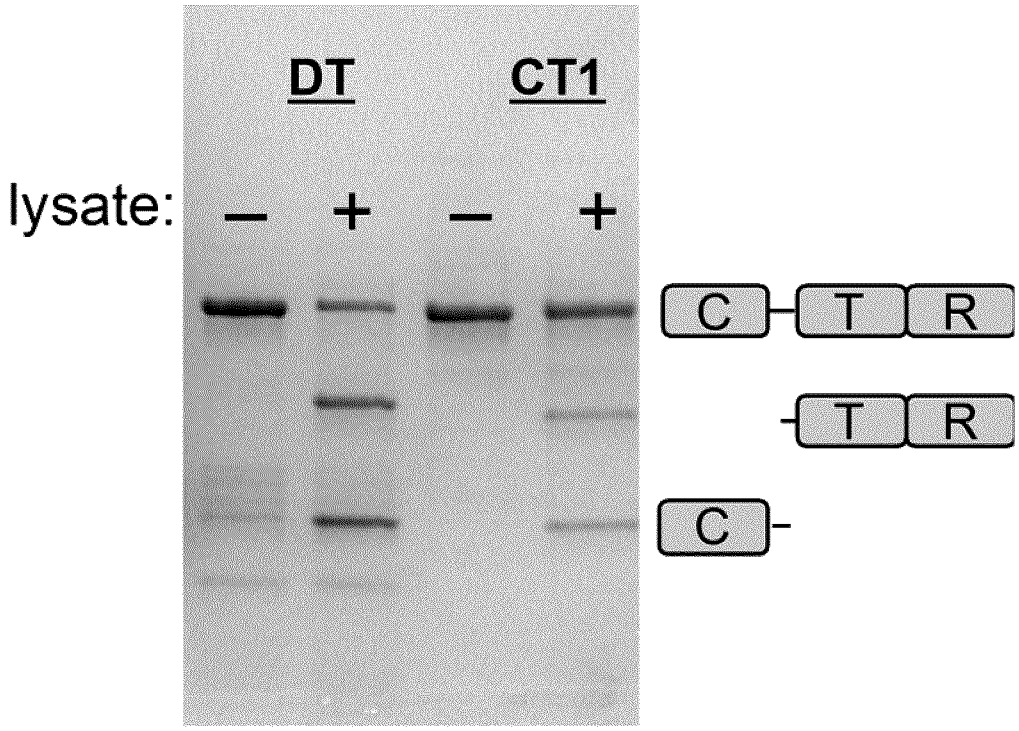



Figure 2C

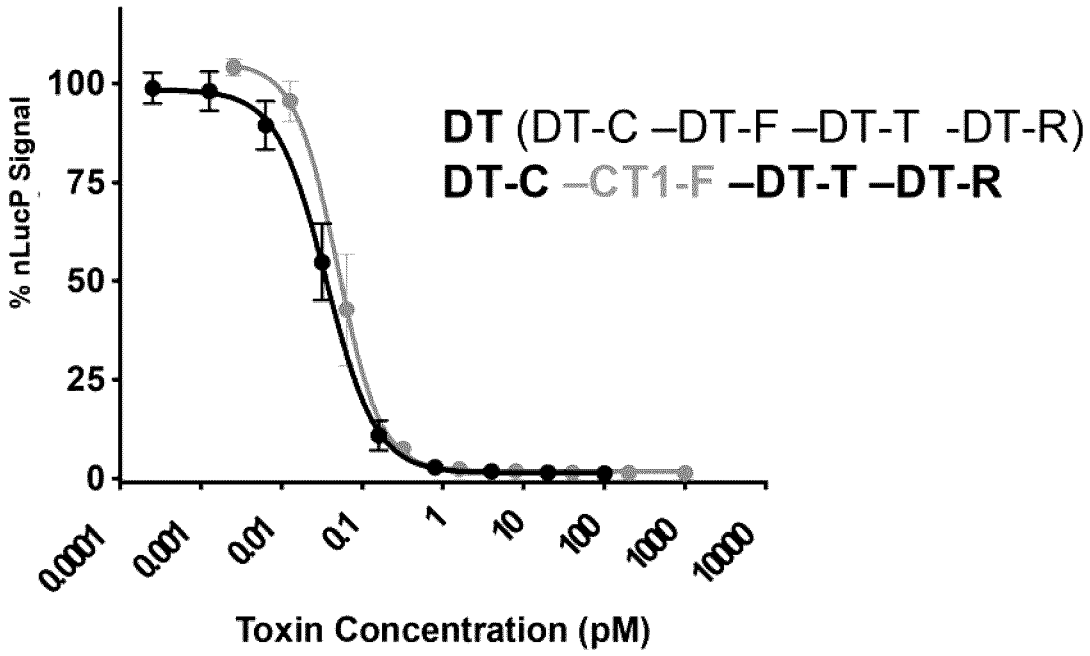

% nLucP Signal
100
75
50
25
0
0.0001
0.001
0.01
0.1
1
10
100
1000
10000
Toxin Concentration (pM)
DT (DT-C –DT-F –DT-T -DT-R)
DT-C –CT1-F –DT-T –DT-R

Figure 3A

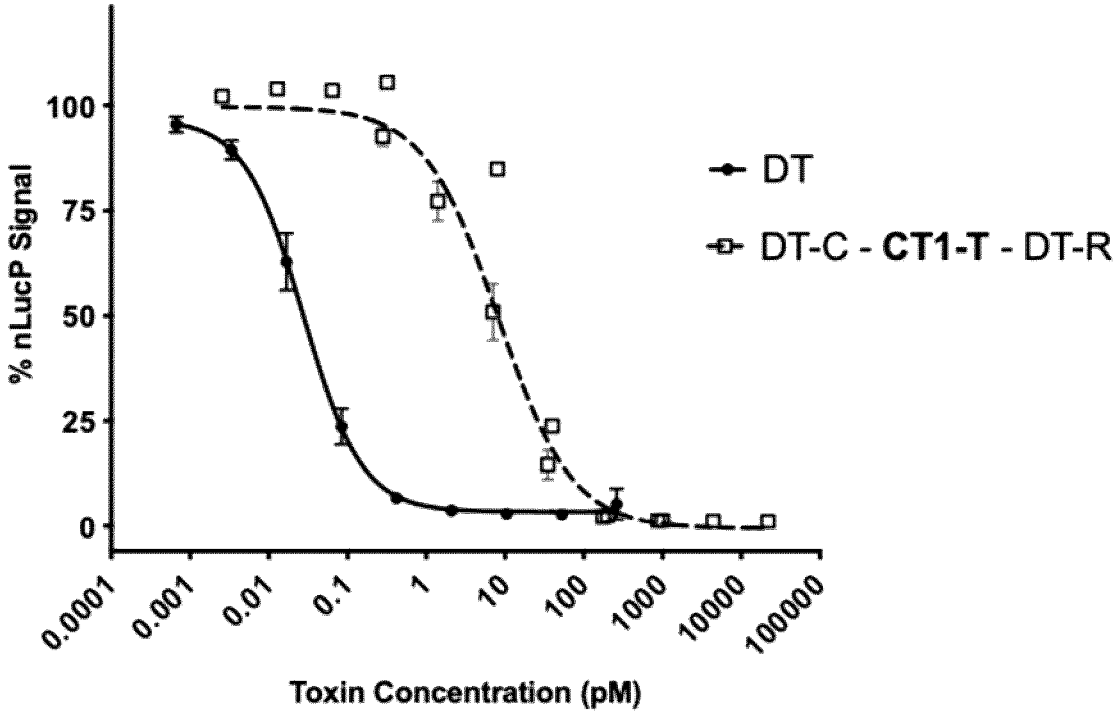

% nLucP Signal
100
75
50
25
0
0.0001
0.001
0.01
0.1
1
10
100
1000
10000
100000
Toxin Concentration (pM)
DT
DT-C - CT1-T - DT-R

Figure 3B

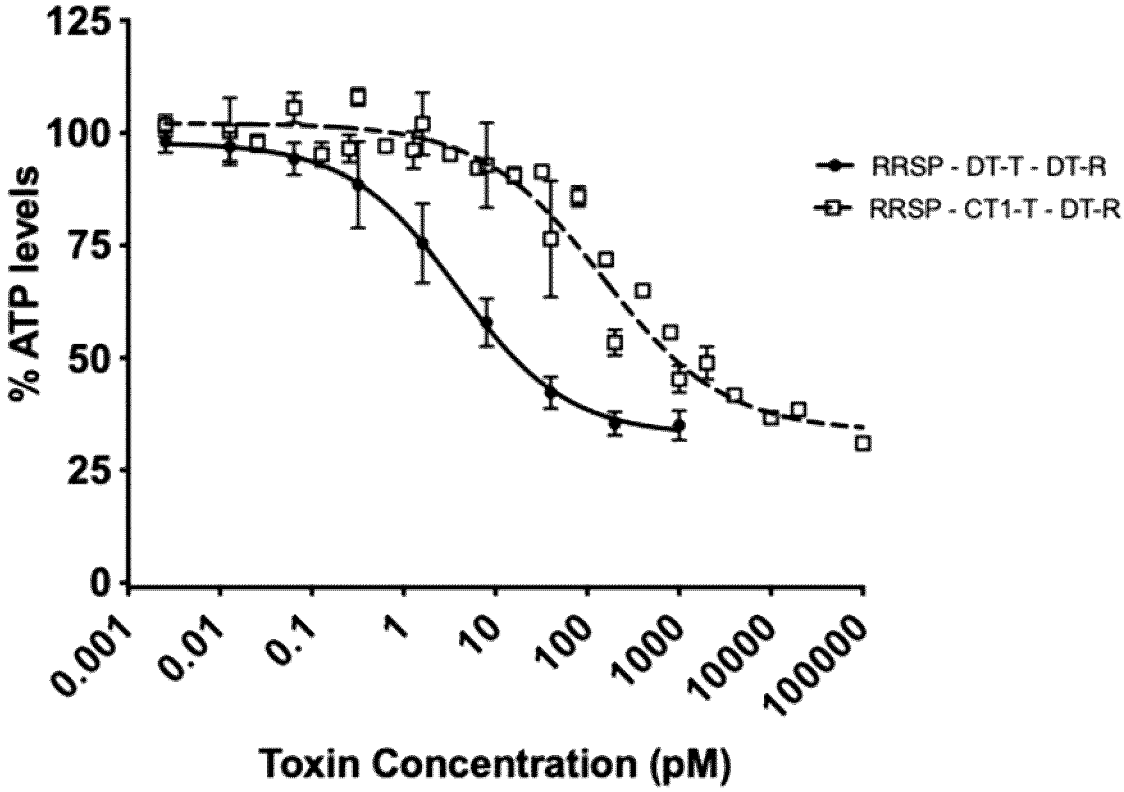

125
100
75
50
25
0
% ATP levels
0.001
0.01
0.1
1
10
100
1000
10000
100000
Toxin Concentration (pM)
RRSP - DT-T - DT-R
RRSP - CT1-T - DT-R

Figure 3C

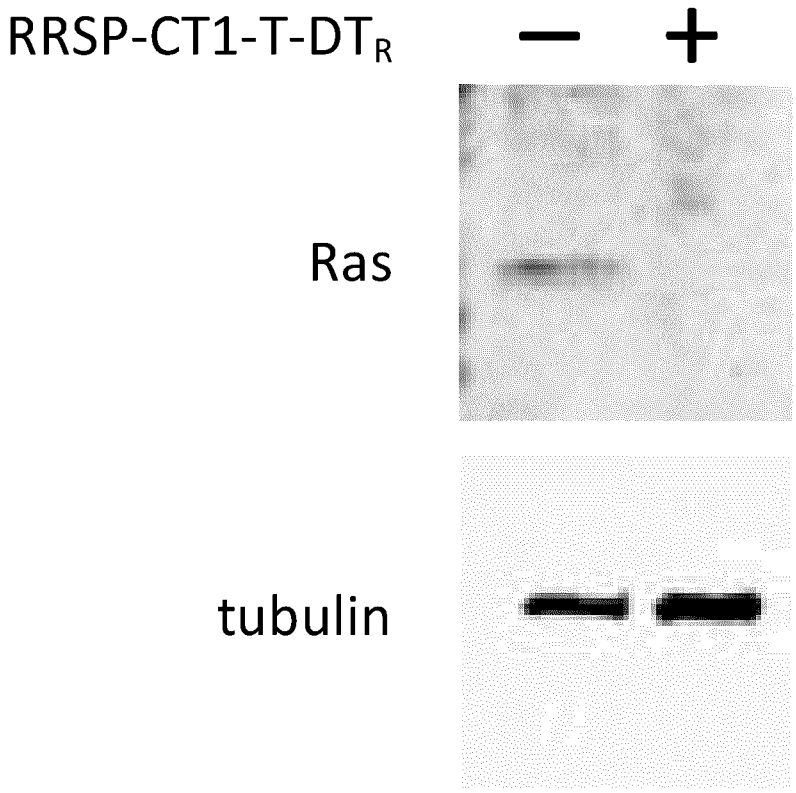



**Figure 3D**

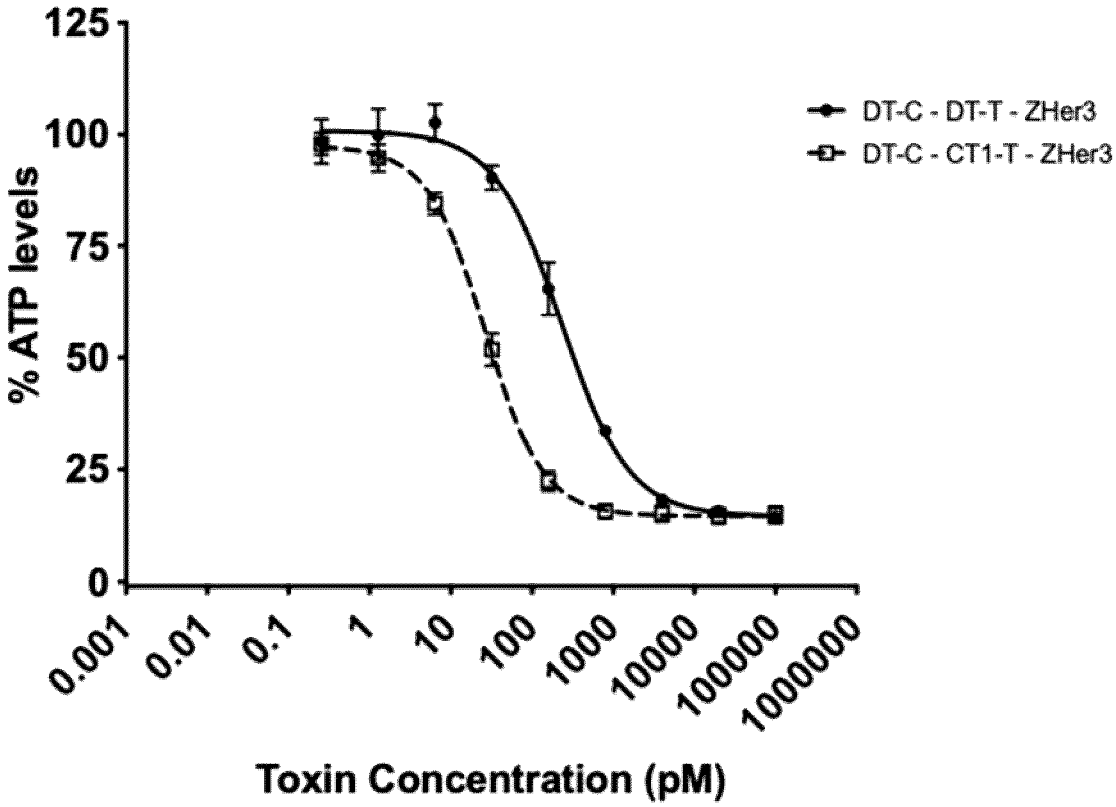

125
100
75
50
25
0
% ATP levels
0.001
0.01
0.1
1
10
100
1000
10000
100000
1000000
Toxin Concentration (pM)
DT-C - DT-T - ZHer3
DT-C - CT1-T - ZHer3

Figure 3E

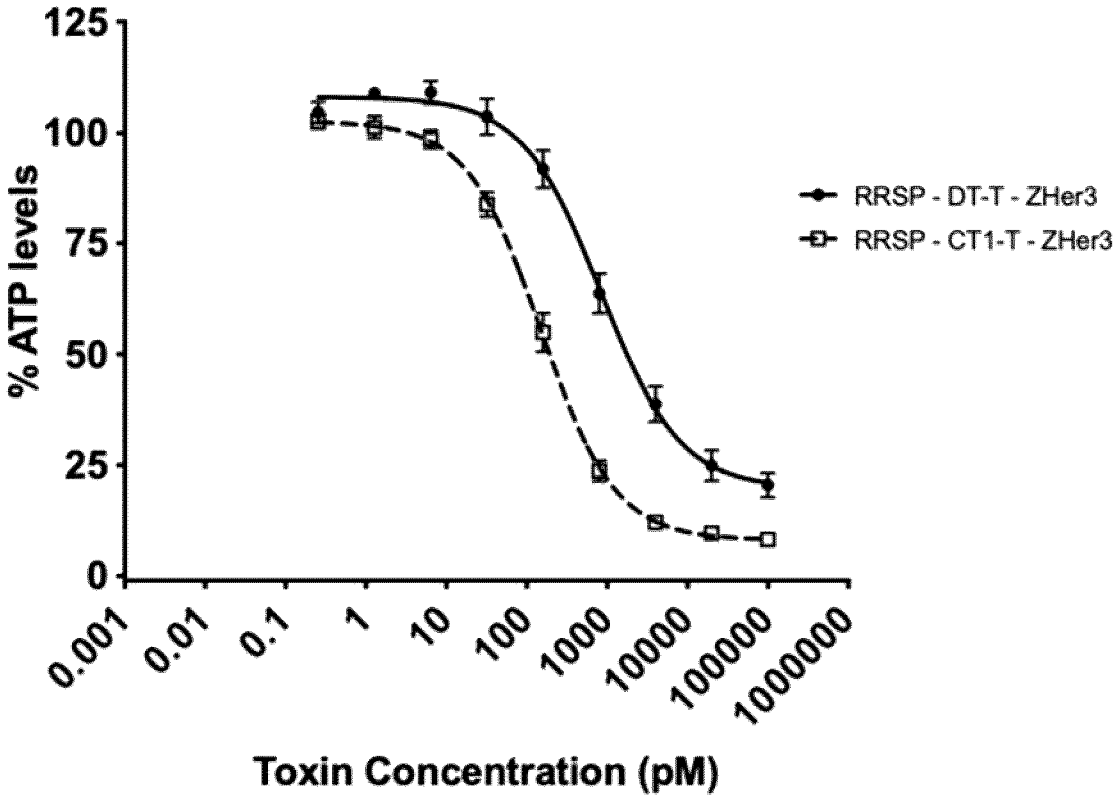

125
100
75
50
25
0
% ATP levels
0.001
0.01
0.1
1
10
100
1000
10000
100000
1000000
Toxin Concentration (pM)
RRSP - DT-T - ZHer3
RRSP - CT1-T - ZHer3

Figure 3F

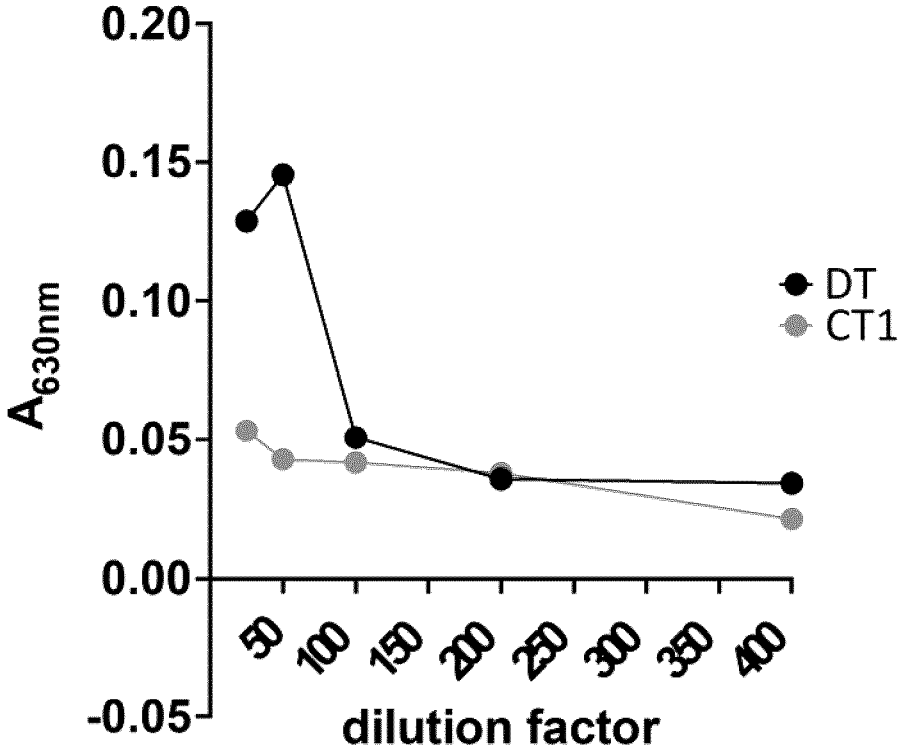

0.20
0.15
0.10
0.05
0.00
-0.05
$A_{630nm}$
50
100
150
200
250
300
350
400
dilution factor
DT
CT1

Figure 4A

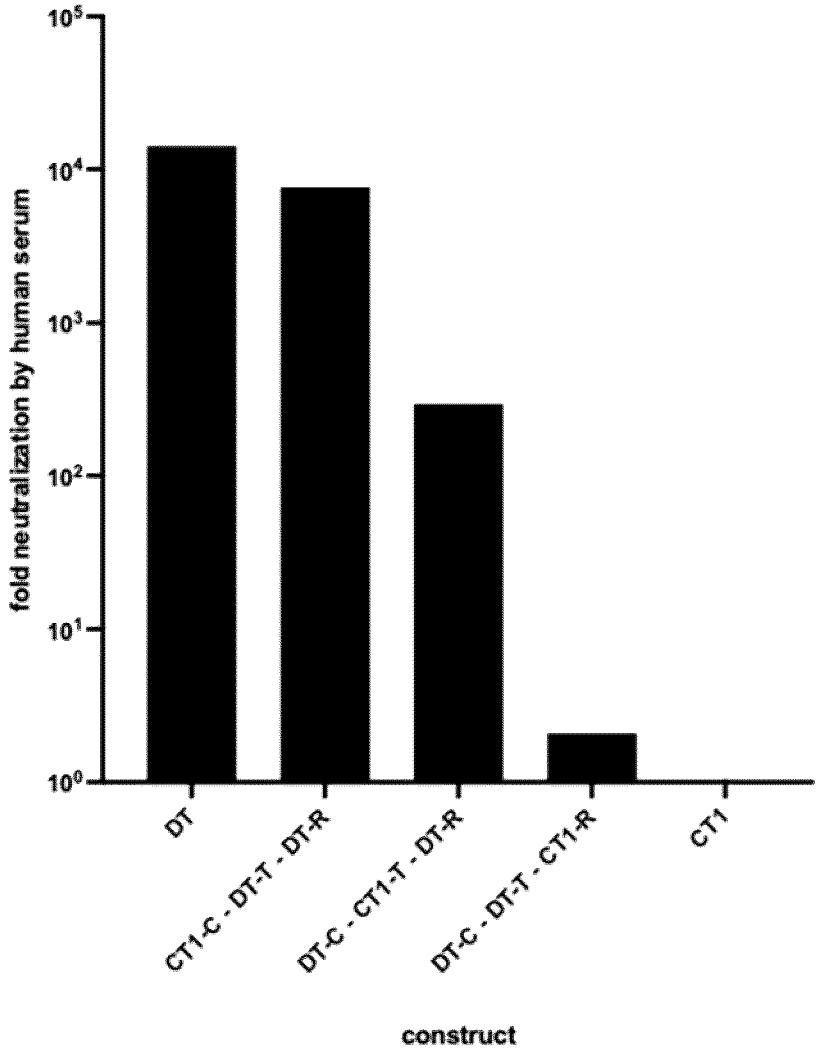

fold neutralization by human serum
$10^5$
$10^4$
$10^3$
$10^2$
$10^1$
$10^0$
DT
CT1-C - DT-T - DT-R
DT-C - CT1-T - DT-R
DT-C - DT-T - CT1-R
CT1
construct

Figure 4B

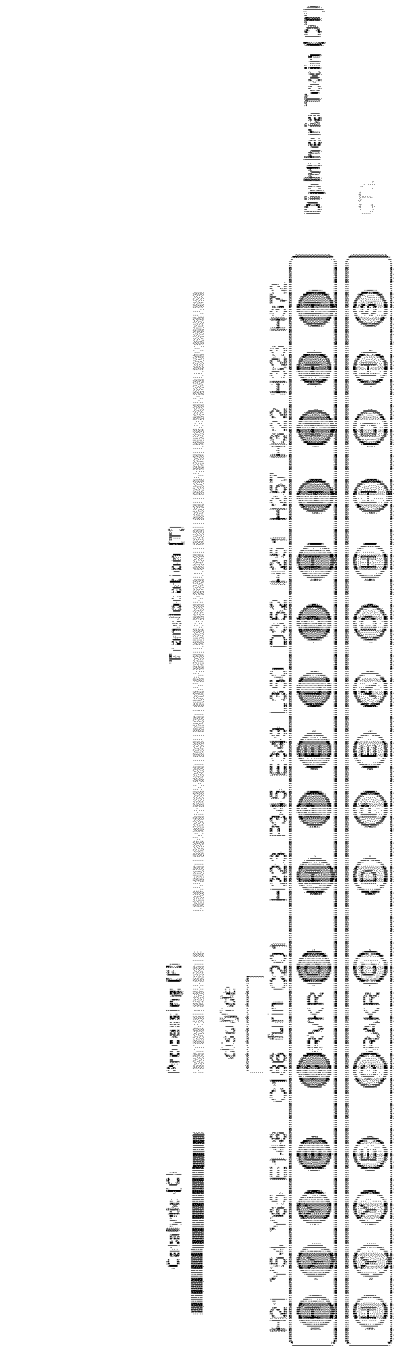

Translocation (T)
Processing (P)
Catalytic (C)
disulfide
H21 Y54 Y65 E148 C186 turn C201 H223 P345 E349 L350 D352 H251 H257 H302 H323 H372
R V K R
R A K R
Diphtheria Toxin (DT)
CT1

Figure 5

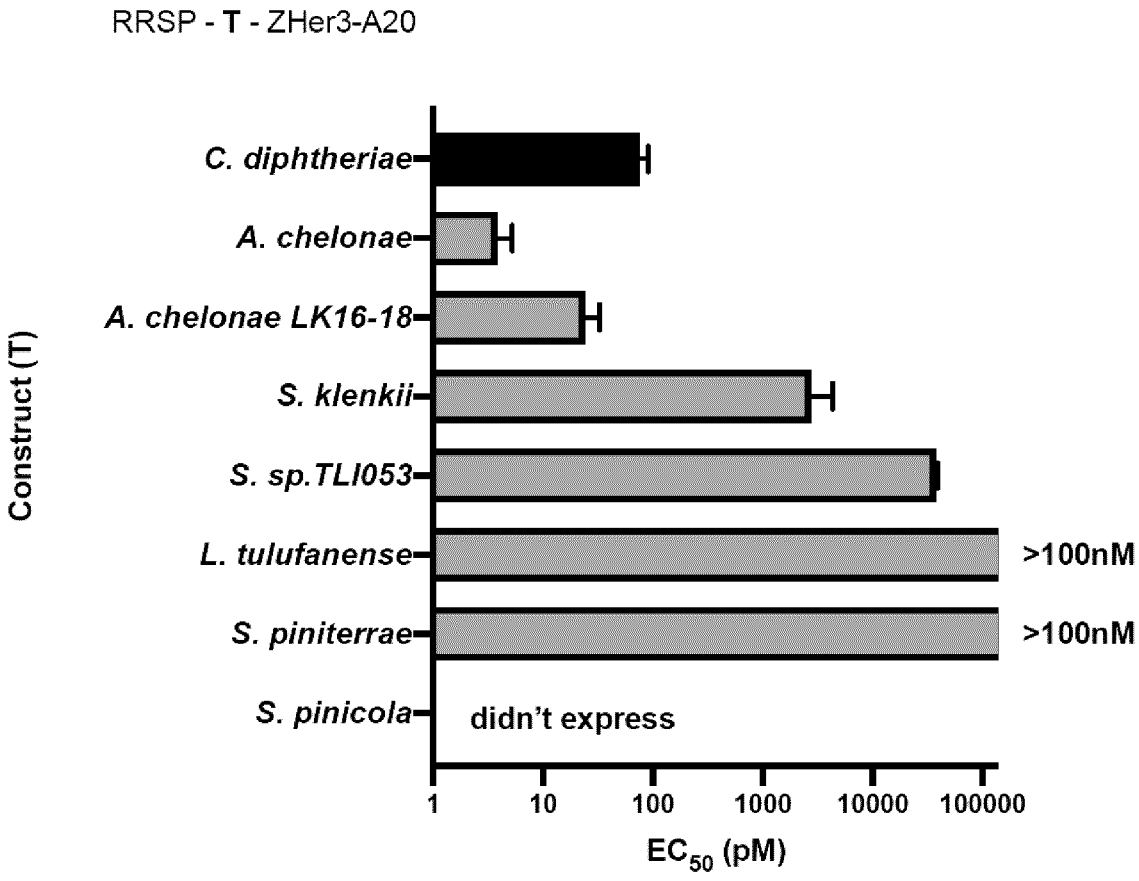

RRSP - T - ZHer3-A20
Construct (T)
C. diphtheriae
A. chelonae
A. chelonae LK16-18
S. klenkii
S. sp.TLI053
L. tulufanense
S. piniterrae
S. pinicola
>100nM
>100nM
didn't express
1
10
100
1000
10000
100000
EC50 (pM)

Figure 6A

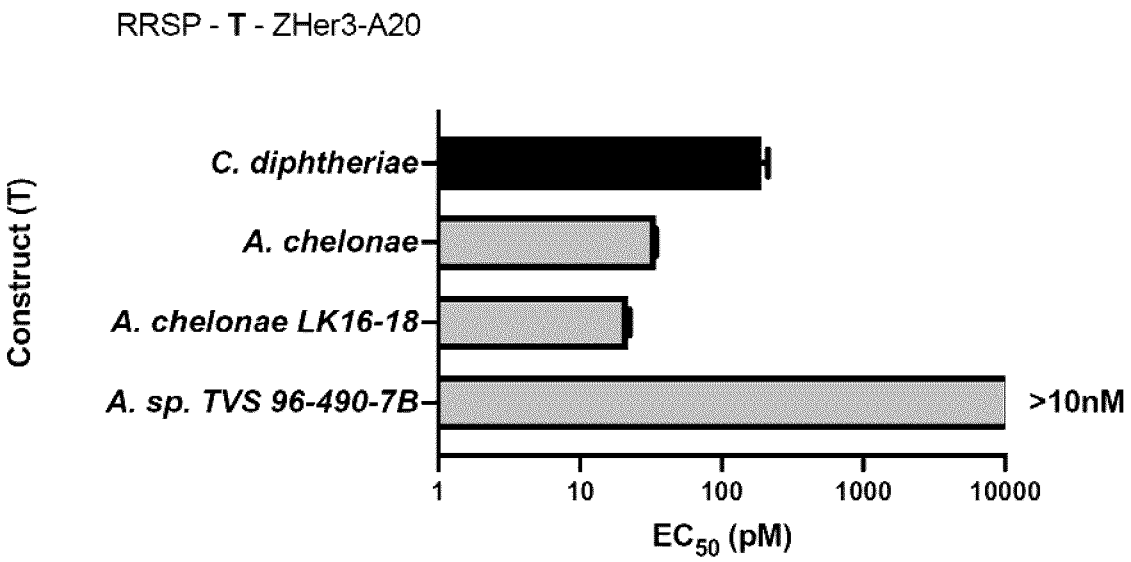

RRSP - T - ZHer3-A20
Construct (T)
C. diphtheriae
A. chelonae
A. chelonae LK16-18
A. sp. TVS 96-490-7B
>10nM
1
10
100
1000
10000
EC50 (pM)

Figure 6B

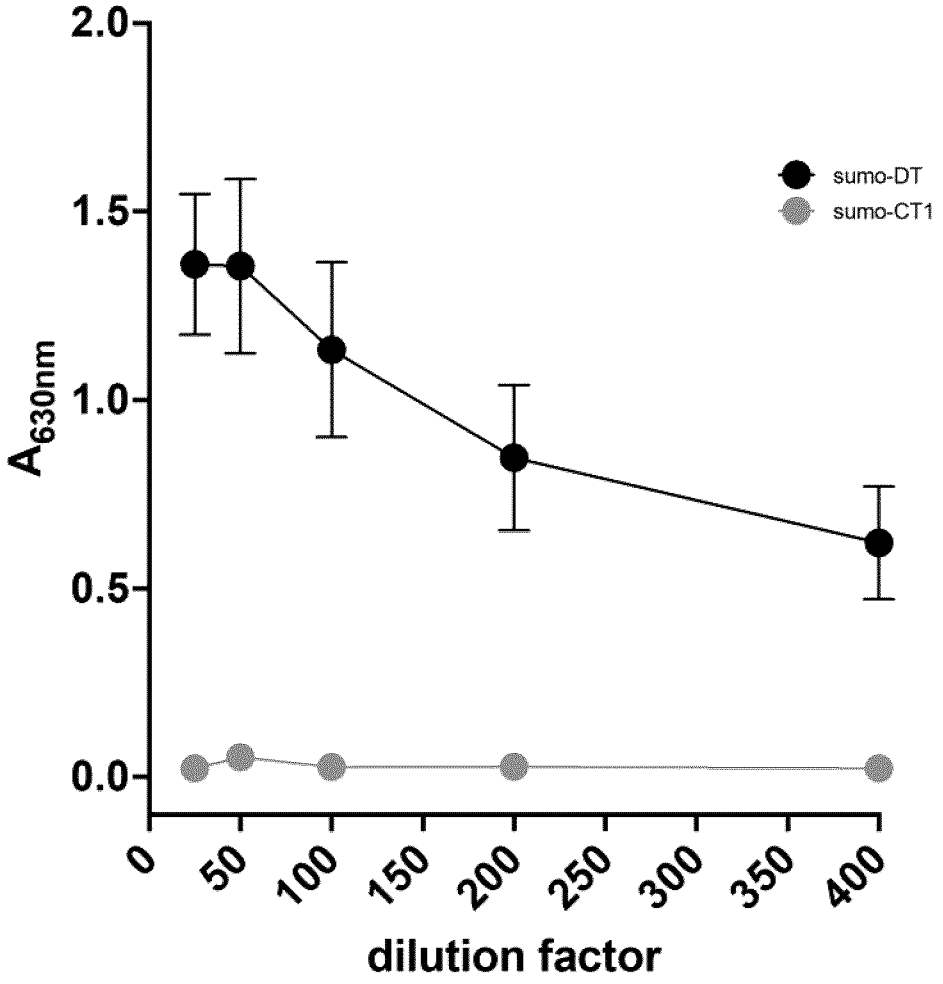

2.0
1.5
1.0
0.5
0.0
$A_{630nm}$
sumo-DT
sumo-CT1
0
50
100
150
200
250
300
350
400
dilution factor

Figure 7A

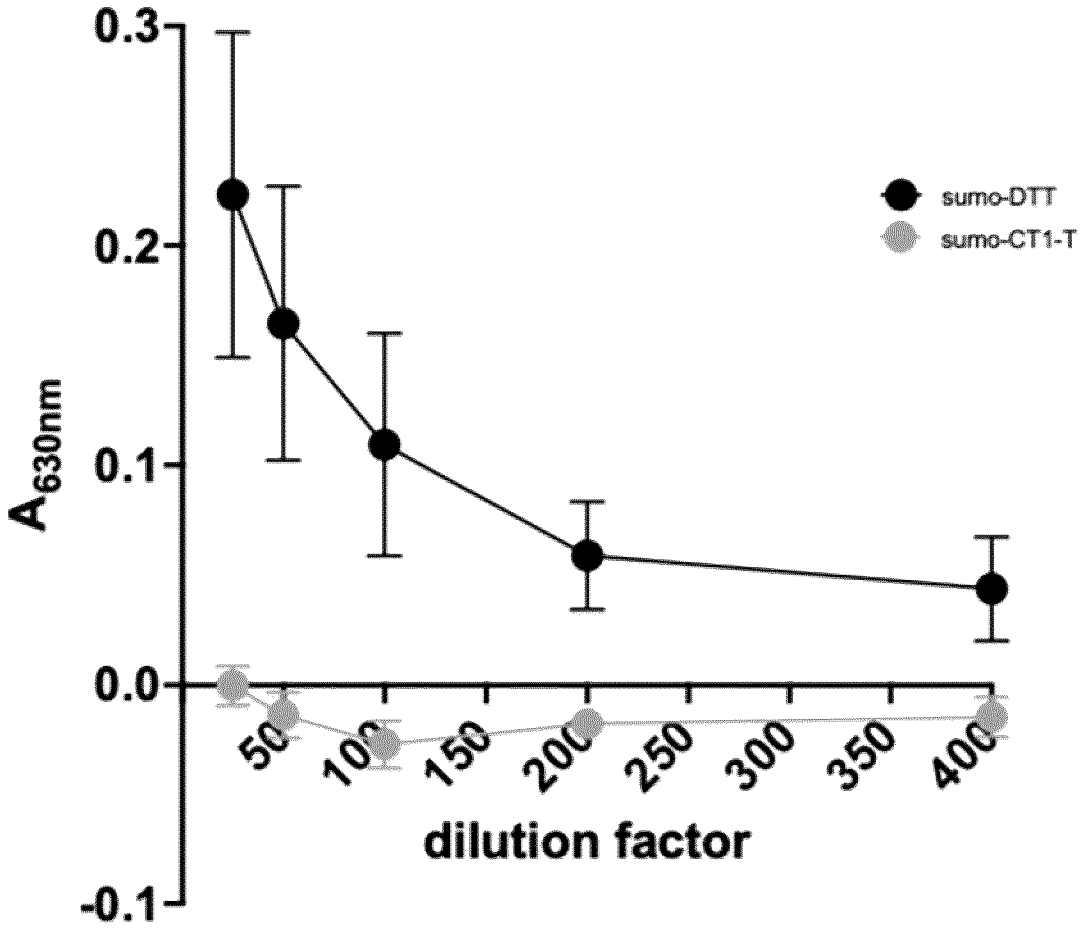

0.3
0.2
0.1
0.0
-0.1
A630nm
sumo-DTT
sumo-CT1-T
50
100
150
200
250
300
350
400
dilution factor

Figure 7B

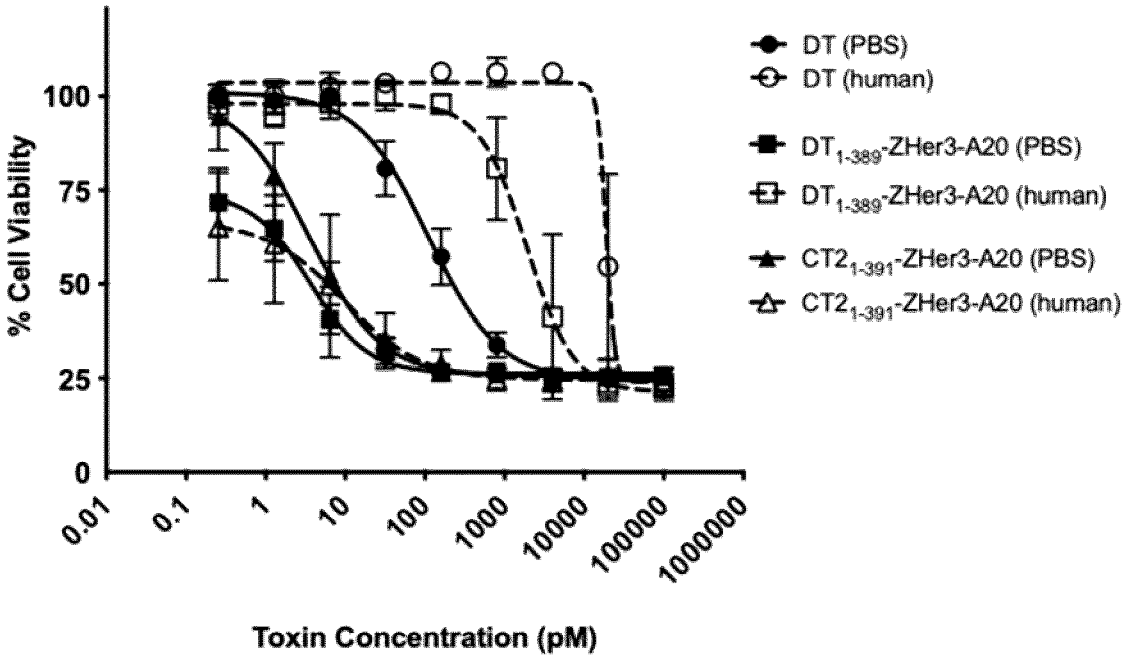

% Cell Viability
100
75
50
25
0
0.01
0.1
1
10
100
1000
10000
100000
1000000
Toxin Concentration (pM)
DT (PBS)
DT (human)
DT1-389-ZHer3-A20 (PBS)
DT1-389-ZHer3-A20 (human)
CT21-391-ZHer3-A20 (PBS)
CT21-391-ZHer3-A20 (human)

Figure 8

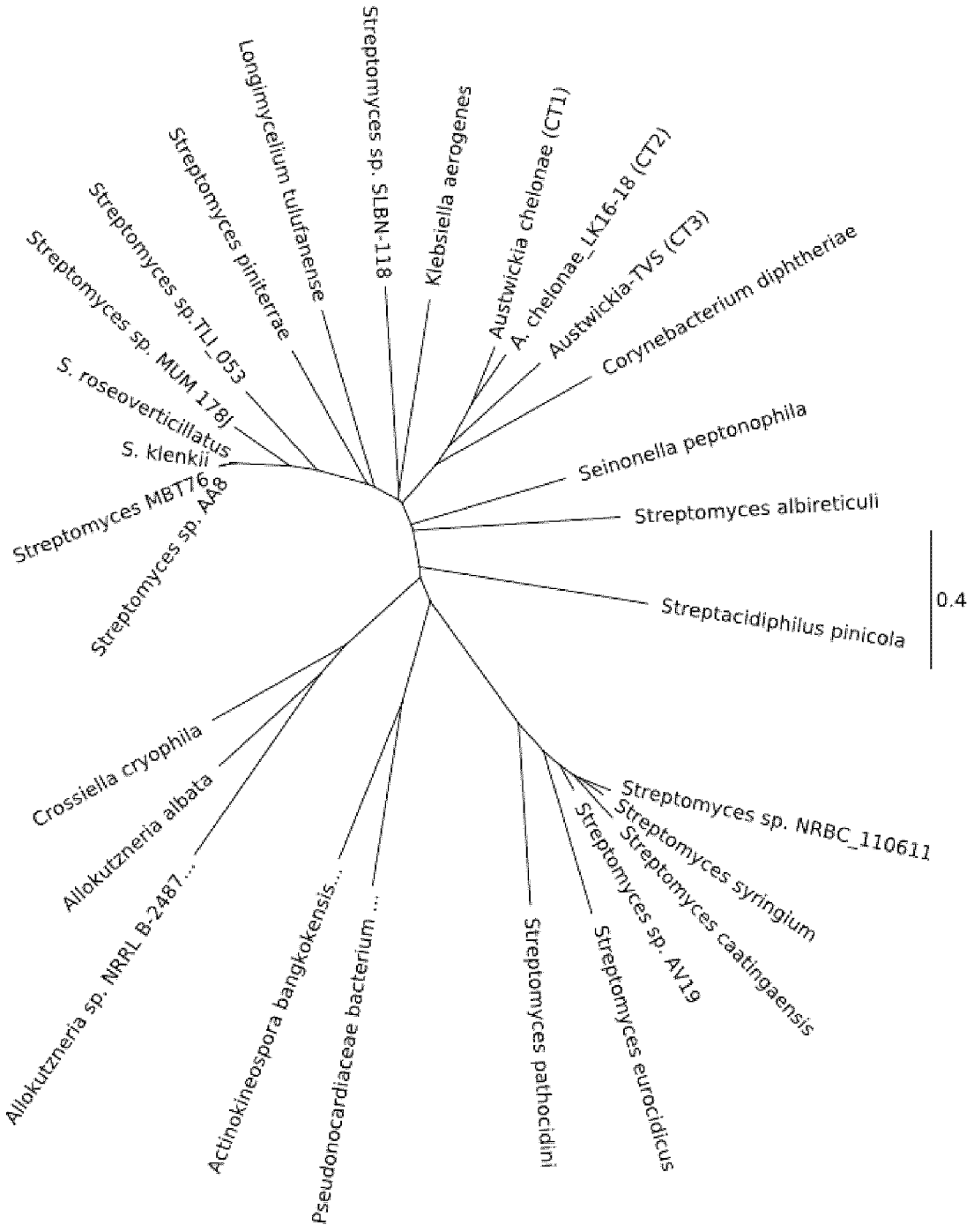

Streptomyces sp. SLBN-118
Klebsiella aerogenes
Austwickia chelonae (CT1)
A. chelonae_LK16-18 (CT2)
Austwickia-TVS (CT3)
Corynebacterium diphtheriae
Seinonella peptonophila
Streptomyces albireticuli
Streptacidiphilus pinicola
0.4
Streptomyces sp. NRBC_110611
Streptomyces syringium
Streptomyces caatingaensis
Streptomyces sp. AV19
Streptomyces eurocidicus
Streptomyces pathocidini
Pseudonocardiaceae bacterium ...
Actinokineospora bangkokensis...
Allokutzneria sp. NRRL B-2487...
Allokutzneria albata
Crossiella cryophila
Streptomyces sp. AA8
Streptomyces MBT76
S. klenkii
S. roseoverticillatus
Streptomyces sp. MUM 178J
Streptomyces sp.TLI_053
Streptomyces piniterrae
Longimycelium tulufanense

**Figure 9**