



(51) International Patent Classification:

A61K 35/28 (2015.01) C12N 5/078 (2010.01)

A61K 48/00 (2006.01)

(21) International Application Number:

PCT/US2019/019117

(22) International Filing Date:

22 February 2019 (22.02.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/633,933 22 February 2018 (22.02.2018) US

62/796,823 25 January 2019 (25.01.2019) US

(71) Applicant: ENTRADA THERAPEUTICS, INC.

[US/US]; 28 State Street, Suite 2303, Boston, Massachusetts 02109 (US).

(72) Inventors; and

(71) Applicants: SETHURAMAN, Natarajan [US/US]; c/o ENTRADA THERAPEUTICS, INC., 28 State Street, Suite 2303, Boston, Massachusetts 02109 (US). RUTH, Jason [US/US]; c/o ENTRADA THERAPEUTICS, INC., 28 State Street, Suite 2303, Boston, Massachusetts 02109 (US). TARTAGLIA, Lou [US/US]; c/o ENTRADA THERAPEUTICS, INC., 28 State Street, Suite 2303, Boston, Massachusetts 02109 (US). PEI, Dehua [US/US]; c/o ENTRADA THERAPEUTICS, INC., 28 State Street, Suite 2303, Boston, Massachusetts 02109 (US). QIAN, Ziqing [CN/US]; c/o ENTRADA THERAPEUTICS, INC., 28 State Street, Suite 2303, Boston, Massachusetts 02109 (US).

(74) Agent: TUSCAN, Michael et al.; c/o Cooley LLP, 1299

Pennsylvania Ave NW, #700, Washington, District of Columbia 20004 (US).

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

AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

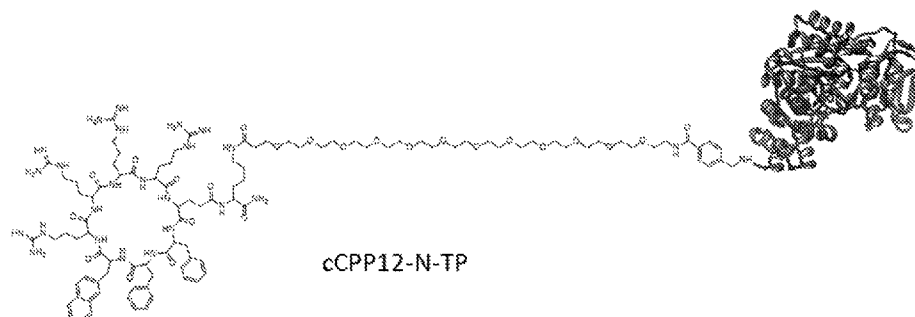
ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG,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

Published:

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

(54) Title: COMPOSITIONS AND METHODS FOR TREATING MITOCHONDRIAL NEUROGASTROINTESTINAL ENCEPHALOPATHY

FIG. 3



(57) Abstract: Disclosed herein are compositions and methods of treating disclosure provides for compounds for use in treating Mitochondrial Neurogastrointestinal Encephalopathy Syndrome (MNGIE). In some embodiments, the compounds have cell penetrating activity and thymidine phosphorylase activity. In certain embodiments, the compounds disclosed herein comprise: a) at least one cell-penetrating peptide (CPP) moiety; and b) a thymidine phosphorylase, or an active fragment or analog thereof (TP), wherein the CPP is coupled, directly or indirectly, to TP.



COMPOSITIONS AND METHODS FOR TREATING MITOCHONDRIAL NEUROGASTROINTESTINAL ENCEPHALOPATHY

CROSS-REFERENCE TO RELATED APPLICATIONS

[1] This Application claims the benefit of U.S. Provisional Application No. 62/633,933, filed February 22, 2018, and U.S. Provisional Application No. 62/796,823, filed January 25, 2019, the entire contents of each of which are hereby incorporated by reference for all purposes.

BACKGROUND

[2] Mitochondrial neurogastrointestinal encephalopathy (MNGIE) disease is a rare, recessive mitochondrial disease that affects several parts of the body, particularly the digestive system and nervous system. Abnormalities of the digestive system are among the most common and severe features of MNGIE disease. Almost all affected people have a condition known as gastrointestinal dysmotility, in which the muscles and nerves of the digestive system do not move food through the digestive tract efficiently. The resulting serious digestive problems (satiety with small amounts of food, dysphagia, nausea, vomiting, abdominal pain, diarrhea, and intestinal blockage) lead to extreme weight loss and reduced muscle mass (cachexia).

[3] MNGIE disease is also characterized by abnormalities of the nervous system. Affected individuals experience peripheral neuropathy, particularly in the hands and feet, as well as other signs and symptoms that can include ptosis, ophthalmoplegia, and hearing loss. Leukoencephalopathy, which is the deterioration of a type of brain tissue known as white matter, is a hallmark of MNGIE disease. These changes in the brain usually do not cause symptoms in people with this disorder.

[4] Mutations in the *TYMP* gene are believed to cause MNGIE disease. This gene provides instructions for making the enzyme thymidine phosphorylase (TP). TP breaks down the DNA building block thymidine into smaller molecules, helping to regulate the level of nucleosides in cells. *TYMP* mutations greatly reduce or eliminate the activity of thymidine phosphorylase. This leads to a toxic level of nucleoside accumulation in the body, which disrupts the usual maintenance and repair of mitochondrial DNA (mDNA). The resulting genetic changes impair the normal function of mitochondria. Although mDNA abnormalities underlie the digestive and neurological problems characteristic of MNGIE disease, how defective mitochondria cause the specific features of the disorder is still under investigation.

[5] For those afflicted with MNGIE, treatment options remain limited. Stem cell transplantation, which suffers from high mortality rates and liver transplantation have been evaluated, but neither affords a general solution. Therefore, management of the disease is primarily through supportive care of the various symptoms and associated ailments. As such, it is clear that new therapies are needed to treat this fatal condition.

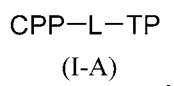
SUMMARY

[6] In various embodiments, the present disclosure provides for compounds for use in treating Mitochondrial Neurogastrointestinal Encephalopathy Syndrome (MNGIE). In embodiments, the compounds have cell penetrating activity and thymidine phosphorylase activity. In certain embodiments, the compounds disclosed herein comprise: a) at least one cell-penetrating peptide (CPP) moiety; and b) a thymidine phosphorylase, or an active fragment or analog thereof (TP), wherein the CPP is coupled, directly or indirectly, to TP.

[7] In some embodiments of the present disclosure, the CPP is conjugated, directly or indirectly, to the TP.

[8] In some embodiments of the present disclosure, the compounds further comprise a linker (L), which conjugates the CPP to TP. In other embodiments, the linker conjugates the CPP to the N-terminus or the C-terminus of the TP. In another embodiment, the linker conjugates the CPP to the N-terminus of the TP. In other embodiments, the linker conjugates the CPP to a side chain of an amino acids in the TP.

[9] In certain embodiments, the compounds disclosed herein have a structure according to Formula I-A:

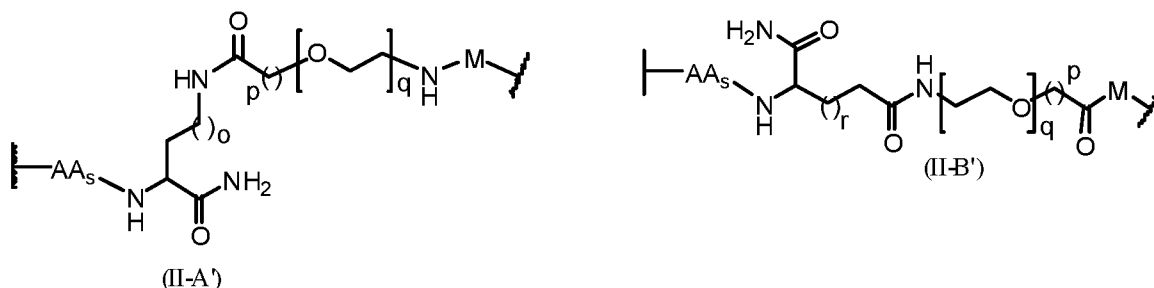


wherein L is a covalently bound to the side chain of an amino acid on the CPP and to the N-terminus of the TP, a side chain of an amino acid in TP, or the C-terminus of the TP. In some embodiments, L is covalently bound to the N-terminus of TP. In other embodiments, L is covalently bound to the C-terminus of TP. In still other embodiments, L is covalently bound to a side chain of an amino acid of TP. In some embodiments, the CPP is a cyclic cell-penetrating peptide (cCPP).

[10] In various embodiments of the present disclosure, L is one or more D or L amino acids, each of which is optionally substituted; alkylene, alkenylene, alkynylene, carbocyclyl, or heterocyclyl, each of which is optionally substituted; or $-(R^1-X-R^2)_z-$, wherein each of R^1 and

R^2 , at each instance, are independently selected from alkylene, alkenylene, alkynylene, carbocyclyl, and heterocyclyl, each X is independently NR^3 , $-NR^3C(O)-$, S, and O, wherein each R^3 is independently selected from H, alkyl, alkenyl, alkynyl, carbocyclyl, and heterocyclyl, each of which is optionally substituted, and z is an integer from 1 to 20; or combinations thereof.

[11] In some embodiments, L has a structure according to Formula II-A' or II-B':



, wherein

M is absent or a group that conjugates L to an amino acid on TP;

AA_s is a side chain or terminus of an amino acid on the CPP;

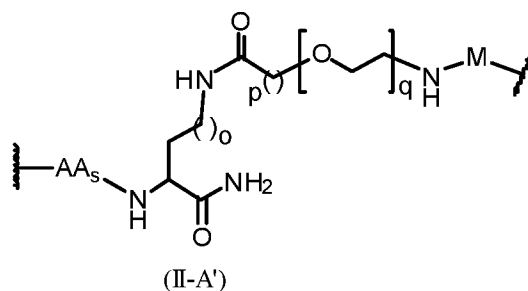
o is an integer from 0 to 10;

p is an integer from 0 to 10;

q is an integer from 1 to 50; and

r is 0 or 1.

[11] In certain embodiments of the present disclosure, L is Formula II-A':



wherein

M is absent or a group that conjugates L to TP;

AA_s is a side chain or terminus of an amino acid on the CPP;

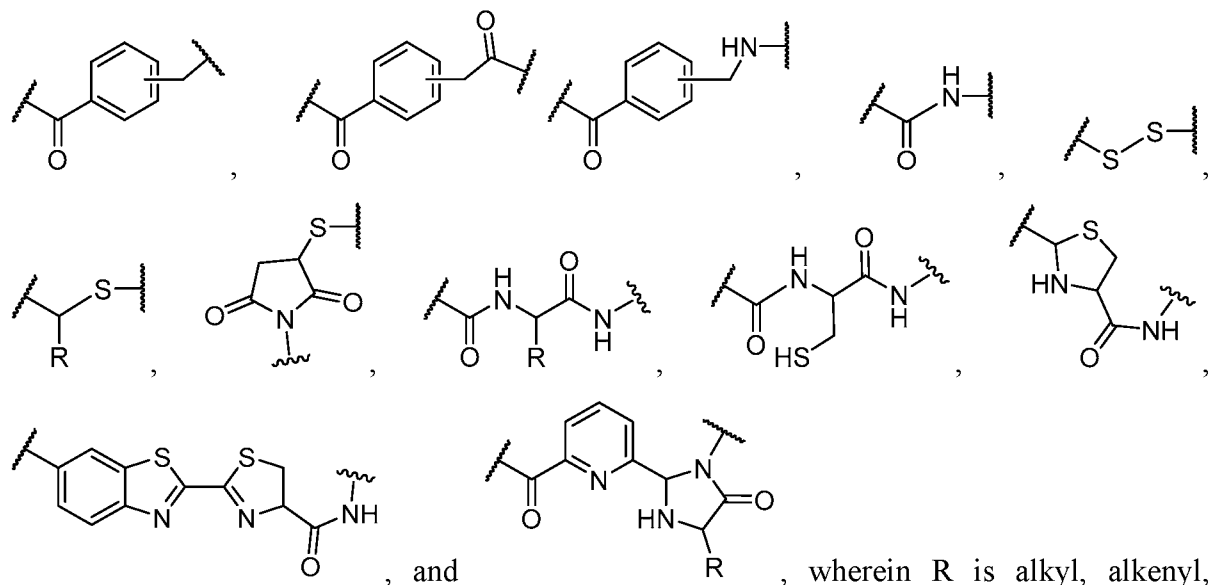
u is 0 or 1;

o is 3;

p is 2; and

q is an integer from 10 to 15;

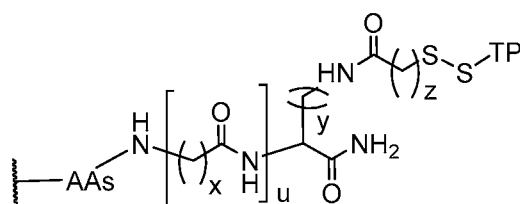
[12] In various embodiments, M is present and comprises an alkylene, alkenylene, alkynylene, carbocyclyl, or heterocyclyl, each of which is optionally substituted. In some embodiments, M is present and selected from the group consisting of:



wherein R is alkyl, alkenyl, alkynyl, carbocyclyl, or heterocyclyl. In a specific embodiment, M is

[13] In some embodiments of the present disclosure, u is 0. In other embodiments, p is 2. In still other embodiments, q is 12. In some embodiments, u is 0, p is 2, and q is 2.

[14] In some embodiments, L is Formula II-C':



II-C'

wherein:

AA_s is a side chain or terminus of an amino acid on the CPP;

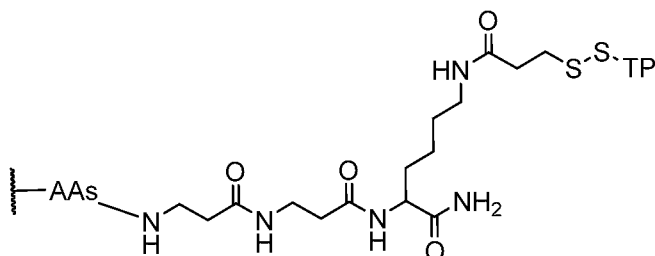
z is an integer from 0 to 10;

y is an integer from 0 to 10;

x is an integer from 0 to 10; and

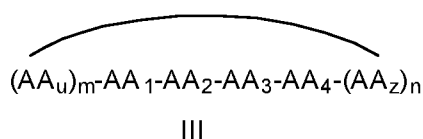
u is an integer from 1 to 50.

[15] In some embodiments, L is



[16] In various embodiments of the present disclosure, L or M is covalently bound to the N-terminus of TP or the C-terminus of TP. In another embodiment, L or M is covalently bound to the N-terminus of TP. In some embodiments, L or M is covalently bound to a side chain of an amino acid in TP (e.g. cysteine).

[17] In some embodiments, the compounds disclosed herein comprise a cCPP which has a sequence comprising Formula III:



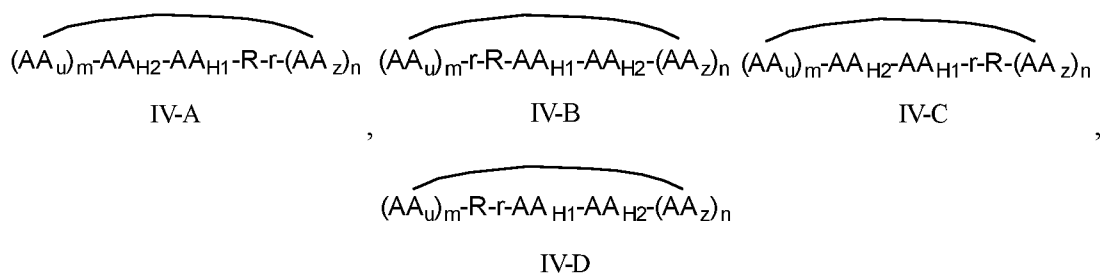
wherein:

each of AA₁, AA₂, AA₃, and AA₄, are independently selected from a D or L amino acid,
 each of AA_u and AA_z, at each instance and when present, are independently selected from a D or L amino acid,
 m and n are independently selected from a number from 0 to 6; and

wherein:

at least two of AA_u, at each instance and when present, AA₁, AA₂, AA₃, AA₄, and AA_z, at each instance and when present, are independently arginine, and
 at least two of AA_u, at each instance and when present, AA₁, AA₂, AA₃, AA₄, and AA_z, at each instance and when present, are independently a hydrophobic amino acid.

[18] In some embodiments, the compound comprises cCPP which has a sequence comprising any of Formula IV-A-D:

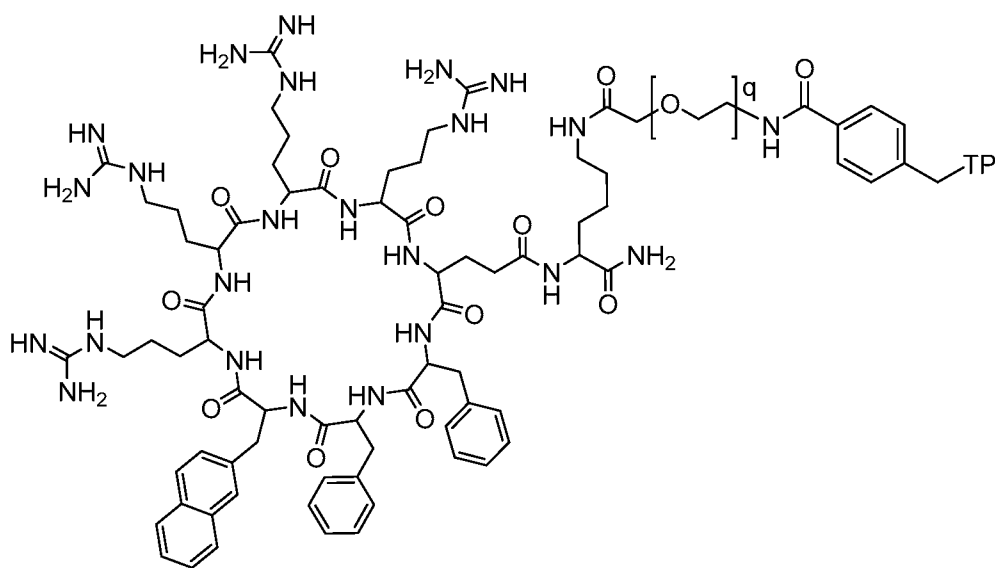


wherein:

each of AA_{H1} and AA_{H2} are independently a D or L hydrophobic amino acid;
 at each instance and when present, each of AA_U and AA_Z are independently a D or L amino acid; and

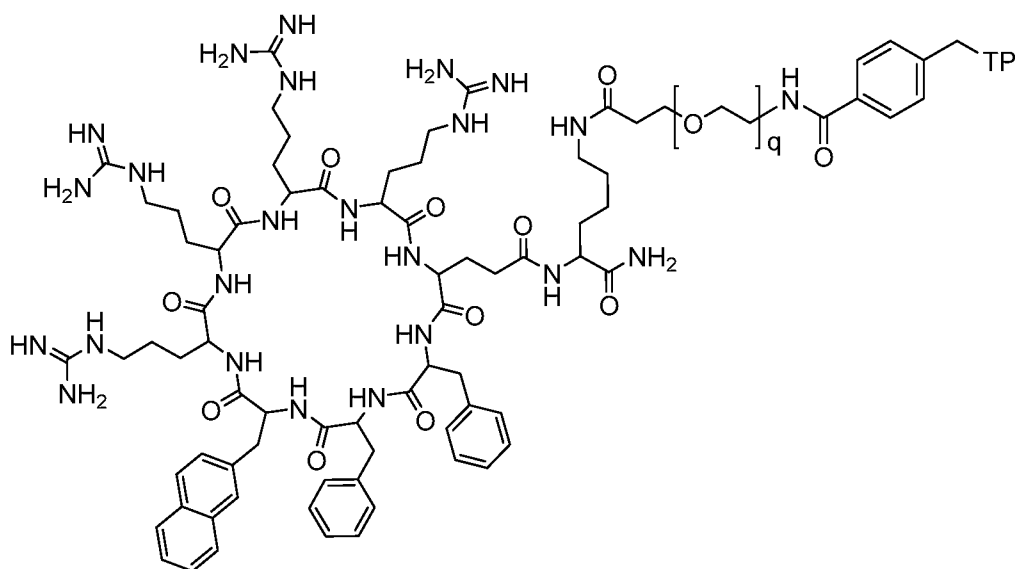
m and n are independently selected from a number from 0 to 6.

[19] In various embodiments of the present disclosure, the compound has a structure according to Formula V-A1 or V-A2:



(V-A1)

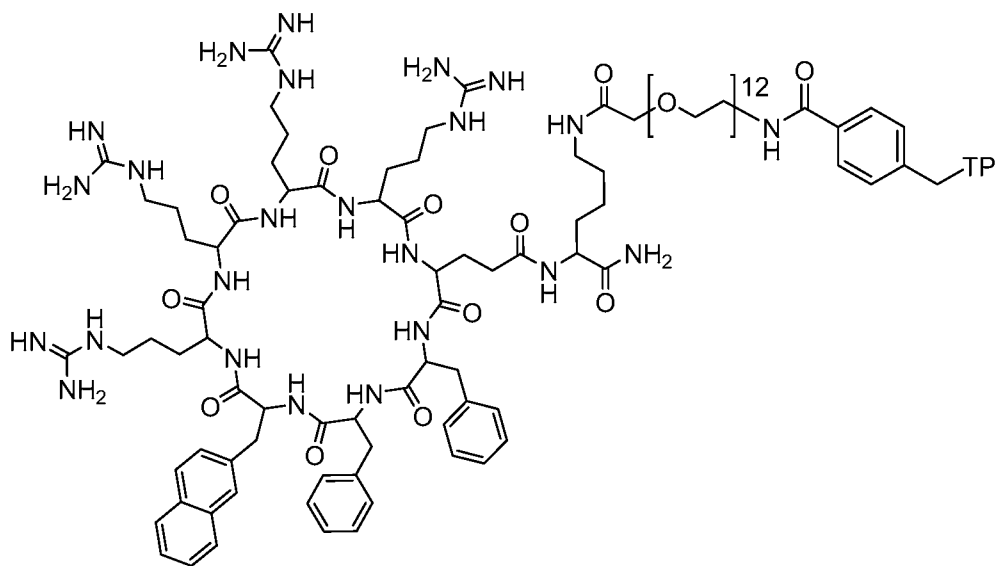
or



(V-A2)

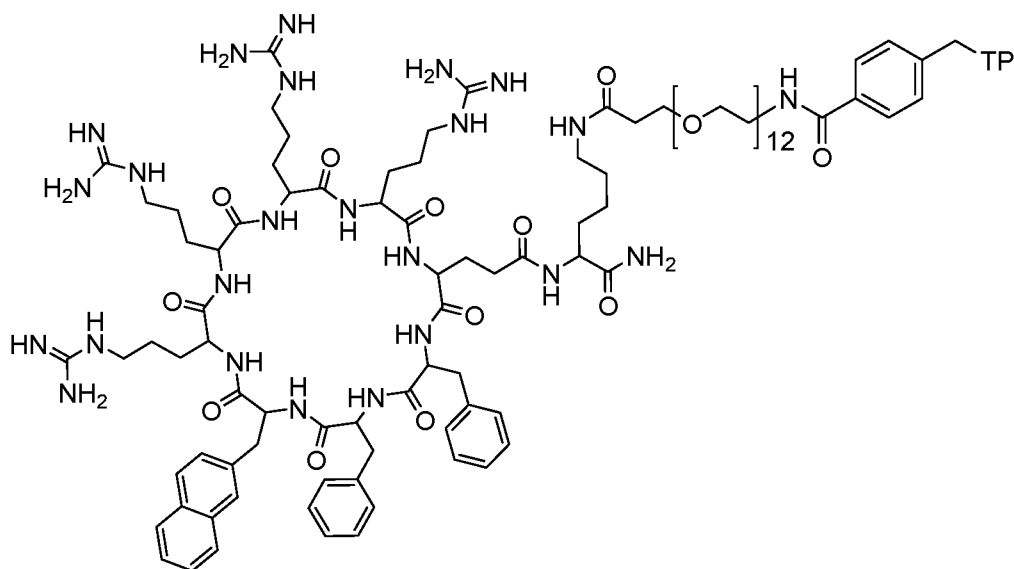
[20] In some embodiments, q is an integer from 1-50. In other embodiments, q is an integer from 10-15. In still other embodiments, q is 12.

[21] In some embodiments, the compounds of the present disclosure have a structure according to Formula V-B1 or V-B2 :



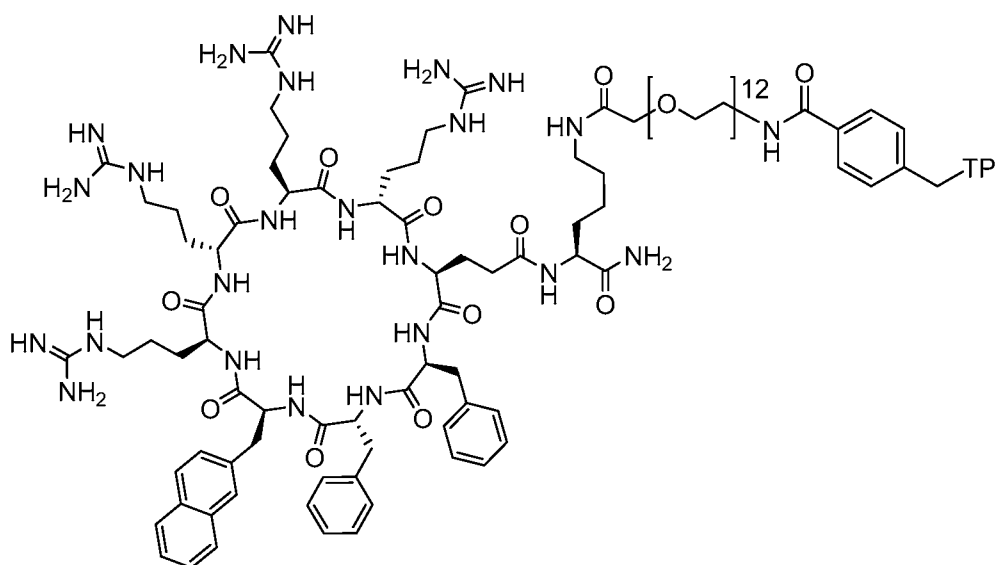
(V-B1)

or



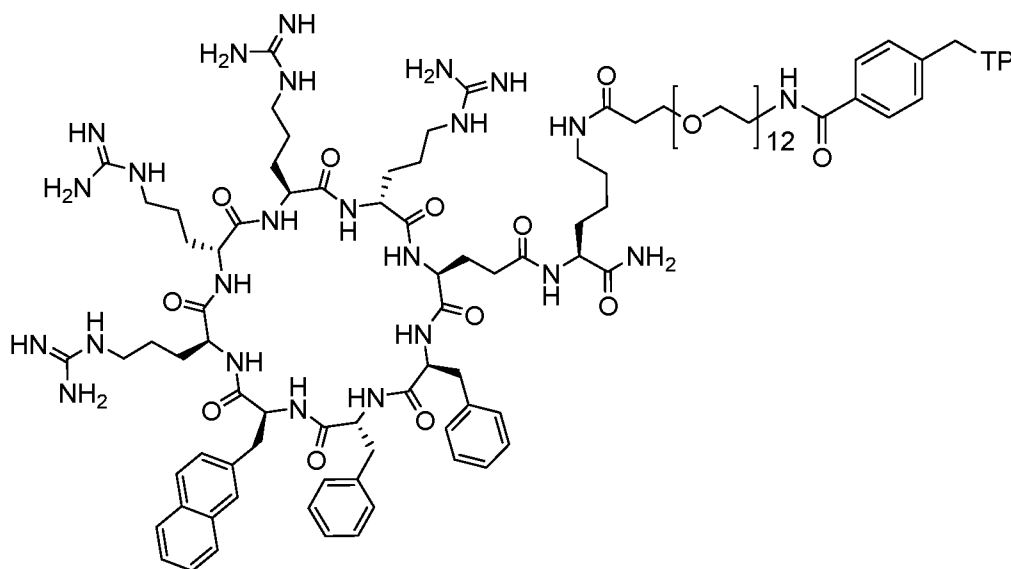
(V-B2)

[22] In other embodiments, the present disclosure provides a compound having the following structure of Formula V-B3 or V-B4:



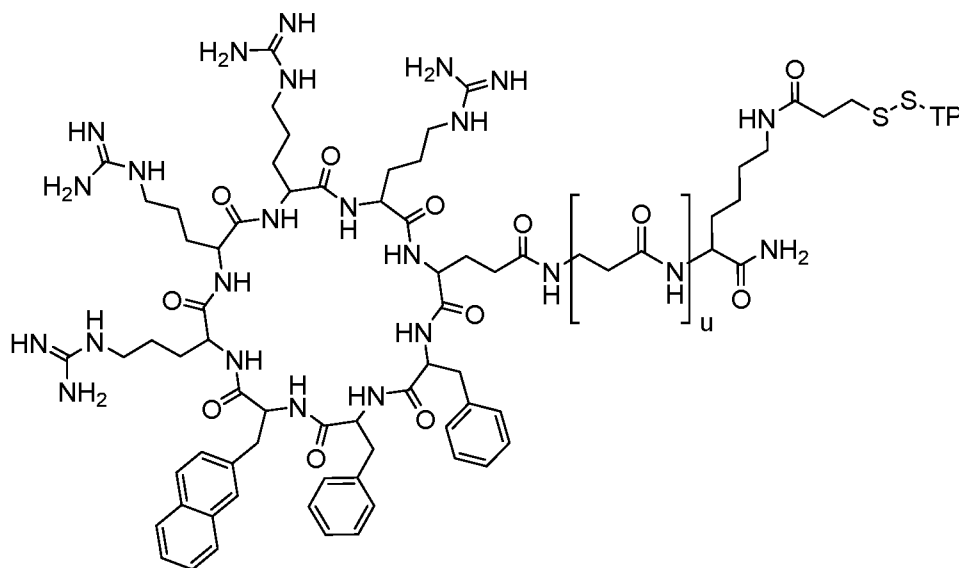
(V-B3)

or



(V-B4)

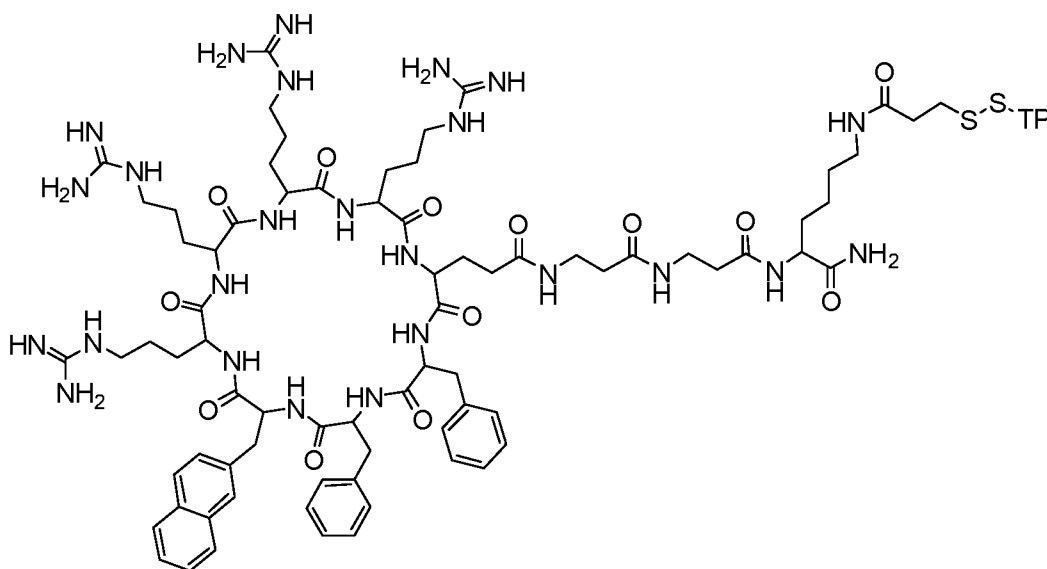
[23] In some embodiments, the present disclosure provides a compound having the following structure of Formula V-A3.



(V-A3)

[24] In some embodiments, u is an integer from 1-50. In other embodiments, u is an integer from 1-5. In still other embodiments, u is 2.

[25] In some embodiments, the compound of Formula V-A3 has the following structure:



[26] In some embodiments, the compounds disclosed herein comprise TP having an amino acid sequence which is at least 85% (e.g., 90%, 95%, or 99%) identical to SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13.

[27] In various embodiments, the present disclosure provides for methods of treating Mitochondrial Neurogastrointestinal Encephalopathy (MNGIE) in a patient in need thereof, comprising administering a compound disclosed herein.

[28] In some embodiments, the present disclosure provides for methods of reducing extracellular and/or intracellular levels of thymidine in a patient in need thereof, comprising administering a compound disclosed herein. In other embodiments, the method is for treating Mitochondrial Neurogastrointestinal Encephalopathy (MNGIE).

[29] In still other embodiments, the present disclosure provides for a cell comprising the compounds disclosed herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[30] **FIG. 1A** is a SDS-PAGE analysis showing the expression and purification of His-tagged thymidine phosphorylase. His-tagged TP11 was expressed from *E. coli* culture using Terrific Broth (TB) at 25 °C overnight induced with 0.25 mM IPTG. From left to right are molecular weight marker (MW), whole cell lysate (WC), supernatant after cell lysis (Sup), flow through from His-Trap column (FT), washing (W), Elutions (E1-E3). Typically obtained 100 mg of His-TP from 1 L of *E. coli* culture

[31] **FIG. 1B** is the SDS-PAGE analysis showing the cleavage of His-tag from His-TP11 by Enterokinase Protease. From left to right: molecular weight marker (MW), before cleavage (lane 1), protease cleavage reaction (lane 2), flow through from His-Trap column (lane 3), elution from His-Trap column (lane 4), and concentrated tag-free TP product (lane 5).

[32] **FIG. 2A** is the SDS-PAGE analysis showing the conjugation and production of CPP12-N-TP11.

[33] **FIG. 2B** is a graph showing the maintenance of enzymatic activity (i.e. enzymatic stability) of TP11 in mouse serum after 2 h and 4 h treatments.

[34] **FIG. 3** is a structural scheme of cCPP12-N-TP produced by reductive amination reaction between TP and cCPP12-PEG₁₂-FBA.

[35] **FIG. 4A** is a graph comparing the enzymatic activity of unconjugated human TP11 (40 nM) with the enzymatic activity of cCPP12-N-TP11 (40 nM).

[36] **FIG. 4B** is a graph comparing the enzymatic activity of CPP12-N-TP11 (40 nM) before (left) and after (right) three cycles of freeze and thaw.

[37] **FIG. 4C** is a graph showing the maintenance of enzymatic activity (i.e. enzymatic stability) of 40 nM cCPP12-N-TP11 in mouse serum or cell growth medium (DMEM + FBS) after 2 h treatment.

[38] **FIG. 5** shows the serum stability of Alexa568-labeled TP11 and Alexa568-labeled cCPP12-N-TP11 in mouse serum after treatments of 0, 2 h, 12 h, or 24 h at 37 °C. No degradation is observed.

[39] **FIG. 6A** shows a Western Blotting analysis comparing the amount of intracellular TP in Hela cells (WT with normal TP levels) to LS174T cells (TP-deficient) pretreated for 6 h with 1) 1 μ M cCPP12-N-TP11; 2) 1 μ M TP11; or 3) control medium. cCPP12-N-TP11 efficiently enters the cell and is enzymatically active.

[40] **FIG. 6B** shows a graph of intracellular TP activity in Hela cells (WT with normal TP levels) to LS174T cells (TP-deficient) pretreated for 6 h with 1) 1 μ M cCPP12-N-TP11; 2) 1 μ M TP11; or 3) control medium. cCPP12-N-TP11 efficiently enters the cell and is enzymatically active.

[41] **FIG. 7** is a Western Blotting analysis showing the dose dependent delivery of TP11 into LS174T cells: 1) LS174T cells treated with media (negative control); 2) LS174T cells incubated with 1 μ M TP11; 3) LS174T cells incubated with 0.1 μ M CPP12-TP11; 4) LS174T

cells incubated with 0.5 μ M CPP12-N-TP11; 5) LS174T cells incubated with 1 μ M CPP12-N-TP11; 6) protein ladder; 7) human liver homogenate (positive control); and 8) 5 ng of fresh TP11.

[42] **FIG. 8** is a graph showing whole cell or cytosolic TP activity delivered into TP-deficient LS174 cells analyzed by TP enzyme activity: 1) LS174T cells treated with media (negative control); 2) LS174T cells incubated with 1 μ M TP11; 3) LS174T cells incubated with 0.1 μ M cCPP12-N-TP11; 4) LS174T cells incubated with 0.5 μ M cCPP12-N-TP11; 5) LS174T cells incubated with 1 μ M CPP12-N-TP11; 6) LS174T cells incubated with 1 μ M cCPP12-N-TP11 and lysed with cytosolic lysis buffer; and 7) 20 nM of fresh cCPP12-N-TP11. The data shown in Fig. 8 is from duplicate experiments.

[43] **FIG. 9** is a graph showing serum thymidine levels in MNGIE mice treated with a CPP12-N-TP11 conjugate disclosed herein compared to a control.

[44] **FIG. 10A** is a graph showing the percent reduction of thymidine concentration in serum on day 27 in MNGIE mice after administration of a cCPP12-N-TP11 conjugate on day 26. **FIG. 10B** is graph showing the detection of an anti-CPP-TP antibody by ELISA in MNGIE mice serum on week four, after once-weekly administration for four weeks. Statistical analysis was performed using a Student's t-test (***) indicates a p-value < 0.0001).

[45] **FIG. 11** shows thymidine levels measured after intravenous injection of cCPP12-N-TP11 conjugates at the following time points: weeks one and two 24 hours after treatment (W1-24hr and W2-24hr, respectively); week three 32 hours after treatment (W3-32hr); and week four 36 hours after treatment (W3-36hr).

[46] **FIG. 12** is a graph showing thymidine phosphorylase levels in the serum of MNGIE mice 24, 32, or 36 hours after treatment with a control, a non-PEGylated cCPP12-N-TP11 conjugate (cCPP12-TP), a fluorescently labeled cCPP-TP conjugate (cCPP12-N-TP11-AF568), and two PEGylated cCPP-TP conjugates (cCPP12-N-TP11-PEG12 and cCPP12-N-TP11-PEG5K).

[47] **FIG. 13.** shows thymidine concentrations (μ M) measured at various time points after intravenous injection of non-PEGylated cCPP-TP conjugate (cCPP12-N-TP11) and a fluorescently labeled cCPP-TP conjugate (cCPP12-N-TP11-AF568) compared to two PEGylated cCPP-TP conjugates (cCPP12-N-TP11-PEG12 and cCPP12-N-TP11-PEG5K).

[48] **FIG. 14** is a graph showing serum thymidine concentrations (μ M) in MNGIE mice after treatment with a PEGylated cCPP-TP conjugate (cCPP12-N-TP11-PEG5K at 16 mpk) compared to a non-PEGylated cCPP-TP conjugate (cCPP12-N-TP11 20 mpk) and a control.

[49] **FIG. 15A** is a graph showing the concentration of cCPP12-N-TP11 (nM) at 0.5 hr, 2 hr, 8 hr, 24 hr, or 48 hr after 20 mpk intravenous injection determined by Western Blotting. **FIG. 15B** is a graph showing the concentration of TP (nM) at 0.5 hr, 2 hr, 8 hr, 24 hr, or 48 hr after 20 mpk intravenous injection as determined by a western blot. # indicates that protein levels were below the limits of detection.

[50] **FIG. 16A** is a bar graph showing the biodistribution of fluorescently labeled cCPP-TP (cCPP12-N-TP11-AF568) 24 hours after intravenous administration as measured by fluorescence in tissue homogenates. **FIG. 16B** is a graph showing the localization of fluorescently labeled cCPP-TP (cCPP12-N-TP11-AF568) in the liver, lungs, and intestines 24 hours after intravenous administration as measured by confocal fluorescence imaging.

[51] **FIG. 17A** is an SDS-PAGE analysis showing the expression and purification of tag-free thymidine phosphorylase. TP16 was expressed from *E. coli* culture using minimal medium as batch medium and yeast extract as fed-batch medium in a PDI loop controlled bioreactor. From left to right are 1. Supernatant of cell lysate, 2. Flow through from the first Phenyl Sepharose chromatograph capture step, 3. Elution with 50% buffer B from Phenyl Sepharose chromatograph, 4. Elution with 75% buffer B from hydrophobicity interaction chromatograph, 5. Elution with 100% buffer B from hydrophobicity interaction chromatograph.

[52] **FIG. 17B** is an SDS-PAGE analysis showing the elution profile from the purification of TP16 using Capto adhere multimodal chromatography.

[53] **FIG. 18** is the structure of cCPP12-SS-SPDP.

[54] **FIG. 19** is an SDS-PAGE analysis showing the cCPP-conjugated TP proteins. 1. TP11 protein, 2. cCPP12-N-TP11 protein, 3. TP16 protein, 4. cCPP12-N-TP16 protein.

[55] **FIG. 20** is the RP-HPLC analysis of TP11, TP16, PEG10K-modified TP11, PEG10K modified TP16, cCPP12-SS-TP11-PEG10K, or cCPP12-SS-TP16-PEG10K.

[56] **FIG. 21** is graph comparing the enzymatic activity of TP11 (right) and TP16 (left).

[57] **FIG. 22** is graph comparing the enzymatic activity of (from right to left) TP11, cCPP12-SS-TP16-PEG10K, cCPP12-SS-TP11-PEG10K, and cCPP12-N-TP16-PEG10K.

[58] **FIG. 23** is a graph shows thymidine (Thd) concentrations (μM) and deoxyuridine (dUrd) concentrations (μM) in the serum measured at 3 or 6 days post intravenous injection of 10 mpk of PEGylated cCPP-TP conjugates: cCPP12-N-TP11-PEG5K, cCPP12-N-TP11-PEG10K, cCPP12-N-TP11-PEG40K linear, or cCPP12-N-TP11-PEG40K branch (W1 data).

[59] **FIG. 24A** is a graph shows thymidine (Thd) concentrations (μM) and deoxyuridine (dUrd) concentrations (μM) in the serum measured at 3 days post intravenous injection of 5

mpk of PEGylated cCPP-TP conjugates: cCPP12-N-TP11-PEG5K, cCPP12-N-TP11-PEG10K, cCPP12-N-TP11-PEG40K linear, or cCPP12-N-TP11-PEG40K branch (W3 data).

[60] **FIG. 24B** is a graph shows thymidine (Thd) concentrations (μM) and deoxyuridine (dUrd) concentrations (μM) in the serum measured at 3 days post intravenous injection of 5 mpk of PEGylated CPP-TP conjugates: cCPP12-N-TP11-PEG5K, cCPP12-N-TP11-PEG10K, cCPP12-N-TP11-PEG40K linear, or cCPP12-N-TP11-PEG40K branch (W4 data).

[61] **FIG. 25** is a graph shows the depletion of thymidine (Thd) concentrations (μM) in the serum measured at 3 days post intravenous injection of 10 mpk (W1) or 5 mpk (W2, W3, and W4) of CPP12-N-TP11-PEG10K over weekly injections in a month.

[62] **FIG. 26A** is a graph shows specific TP activity in the serum measured at 3 days post intravenous injection of 10 mpk of PEGylated CPP-TP conjugates: cCPP12-N-TP11-PEG5K, cCPP12-N-TP11-PEG10K, cCPP12-N-TP11-PEG40K linear, or cCPP12-N-TP11-PEG40K branch (W1 data).

[63] **FIG. 26B** is a graph shows specific TP activity in the serum measured at 3 days post intravenous injection of 5 mpk of PEGylated CPP-TP conjugates: cCPP12-N-TP11-PEG5K, cCPP12-N-TP11-PEG10K, cCPP12-N-TP11-PEG40K linear, or cCPP12-N-TP11-PEG40K branch (W4 data).

[64] **FIG. 27A** is a graph shows thymidine (Thd) concentrations in the serum measured at 5 min, 8 hr, 24 hr, 36 hr, or 48 hr post intravenous injection of 5 mpk of cCPP12-N-TP11, cCPP12-N-TP16, or PBS control.

[65] **FIG. 27B** is a graph shows specific TP activity in the serum measured at 5 min, 8 hr, 24 hr, 36 hr, or 48 hr post intravenous injection of 5 mpk of cCPP12-N-TP11, cCPP12-N-TP16, or PBS control.

[66] **FIG. 28A** is a graph shows thymidine (Thd) concentrations in the serum measured at 5 min, 8 hr, 24 hr, 48 hr, 72 hr, or 96 hr post intravenous injection of 5 mpk of cCPP12-N-TP11-PEG10K, cCPP12-N-TP16-PEG10K, or PBS control.

[67] **FIG. 28B** is a graph shows specific TP activity in the serum measured at 5 min, 8 hr, 24 hr, 48 hr, 72 hr, or 96 hr post intravenous injection of 5 mpk of cCPP12-N-TP11-PEG10K, cCPP12-N-TP16-PEG10K, or PBS control.

[68] **FIG. 29A** is a graph shows thymidine (Thd) concentrations in the serum measured at 5 min, 8 hr, 24 hr, 48 hr, 72 hr, or 96 hr post intravenous injection of 5 mpk of cCPP12-SS-TP11-PEG10K, or cCPP12-SS-TP16-PEG10K.

[69] **FIG. 29B** is a graph shows specific TP activity in the serum at 5 min, 8 hr, 24 hr, 48 hr, 72 hr, or 96 hr post intravenous injection of 5 mpk of cCPP12-SS-TP11-PEG10K, or cCPP12-SS-TP16-PEG10K.

[70] **Fig. 30** is an SDS-PAGE analysis showing the Fc-TP16 and CPP conjugated Fc-TP16. Lane 1. Purified Fc-TP16 (SEQ ID NO. 7 in Table 6), and lane 2. cCPP12-N-Fc-TP16 conjugated protein.

[71] **Fig. 31** is a is graph comparing the enzymatic activity of Fc-TP16 (left), cCPP12-N-FcTP16 (middle), and unconjugated TP11 (right).

[72] **FIG. 32** is a graph shows thymidine (Thd) concentrations in the serum measured at 5 min, 24 hr, 48 hr, 72 hr, 120 hr, or 156 hr post intravenous injection of 5 mpk of Fc-TP16, or cCPP12-N-Fc-TP16.

[73] **FIG. 33** is a graph shows specific TP activity in the serum at 5 min, 24 hr, 48 hr, 72 hr, 120 hr, or 156 hr post intravenous injection of 5 mpk of Fc-TP16, or cCPP12-N-Fc-TP16.

[74] **FIG. 34** is a graph showing cytosolic TP activity delivered into TP-deficient LS174 cells as analyzed by TP enzyme activity: 1) LS174T cells treated with media (negative control); 2) LS174T cells incubated with 1 micromolar TP11; 3) LS174T cells incubated with 1 micromolar cCPP12-N-TP11; 4) LS174T cells incubated with 1 micromolar cCPP12-N-TP11-PEG10K; 5) LS174T cells incubated with 1 micromolar cCPP12-N-TP16-PEG10K; (6) LS174T cells incubated with 1 micromolar cCPP12-SS-TP11-PEG10K; and (7) LS174T cells incubated with 1 micromolar cCPP12-SS-TP16-PEG10K.

[75] **FIG. 35** is the SDS-PAGE analysis showing the refolding process of TP16. From left to right: molecular weight marker (lane 1), dissolved inclusion body (lane 2), refolded protein after rapid dilution (lane 3), flow through from Q sepharase chromatography.

[76] **Fig. 36** is a graph comparing the enzymatic activity of TP16 purified from soluble fraction (left) and TP16 obtained from refolding process (right).

DETAILED DESCRIPTION

Definitions

[77] The term “pharmaceutically acceptable” means suitable for use in contact with the tissues of humans and animals without undue toxicity, irritation, allergic response, and the like, commensurate with a reasonable benefit/risk ratio, and effective for their intended use within the scope of sound medical judgment.

[78] The term “pharmaceutically acceptable salts” include those obtained by reacting the active compound functioning as a base, with an inorganic or organic acid to form a salt, for example, salts of hydrochloric acid, sulfuric acid, phosphoric acid, methanesulfonic acid, camphorsulfonic acid, oxalic acid, maleic acid, succinic acid, citric acid, formic acid, hydrobromic acid, benzoic acid, tartaric acid, fumaric acid, salicylic acid, mandelic acid, carbonic acid, etc. Those skilled in the art will further recognize that acid addition salts may be prepared by reaction of the compounds with the appropriate inorganic or organic acid via any of a number of known methods. The term “pharmaceutically acceptable salts” also includes those obtained by reacting the active compound functioning as an acid, with an inorganic or organic base to form a salt, for example salts of ethylenediamine, N-methyl-glucamine, lysine, arginine, ornithine, choline, N,N'-dibenzylethylenediamine, chlorprocaine, diethanolamine, procaine, N-benzylphenethylamine, diethylamine, piperazine, tris-(hydroxymethyl)-aminomethane, tetramethylammonium hydroxide, triethylamine, dibenzylamine, ephenamine, dehydroabietylamine, N-ethylpiperidine, benzylamine, tetramethylammonium, tetraethylammonium, methylamine, dimethylamine, trimethylamine, ethylamine, basic amino acids, and the like. Non limiting examples of inorganic or metal salts include lithium, sodium, calcium, potassium, magnesium salts and the like.

[79] As used herein, “treat,” “treating,” “treatment” and variants thereof, refers to any administration of thymidine phosphorylase (TP) that partially or completely alleviates, ameliorates, relieves, inhibits, delays onset of, reduces severity of, and/or reduces incidence of one or more symptoms or features Mitochondrial Neurogastrointestinal Encephalopathy (MNGIE) as described herein.

[80] As used herein, “therapeutically effective” refers to an amount of TP which confers a therapeutic effect on a patient. In some embodiments, the therapeutically effective amount is an amount sufficient to treat MNGIE.

[81] As used herein, “cell penetrating peptide” or “CPP” refers to any peptide which is capable of penetrating a cell membrane. In some embodiments, the cyclic cell penetrating peptide is also capable of directing a protein (e.g., TP) to penetrate the membrane of a cell. In some embodiments, the cell penetrating peptide is a cyclic cell-penetrating peptide (cCPP). In some embodiments, the CPP delivers the protein to the cytosol of the cell. Without being bound by theory, the CPPs (e.g., cCPPs) deliver the cargo to the cytosol by enabling escape of the CPP-TP conjugate from endosomes.

[82] As used herein, “linker” or “L” refers to a moiety which that covalently bonds two or more moieties (e.g., a cCPP and TP). In some embodiments, the linker can be natural or non-natural amino acid or polypeptide. In other embodiments, the linker is a synthetic compound containing two or more appropriate functional groups suitable to bind a CPP and TP, to thereby form the compounds disclosed herein. In yet another embodiment, the linker comprises an M moiety to thereby conjugate the CPP to the TP. For example, in some embodiments, the cCPP may be covalently bound to TP via a linker.

[83] As used herein, “polypeptide” refers to a string of at least two amino acids attached to one another by a peptide bond. There is no upper limit to the number of amino acids that can be included in a polypeptide. Further, polypeptides may include non-natural amino acids, amino acid analogs, or other synthetic molecules that are capable of integrating into a polypeptide.

[84] As used herein, the “sequence identity” refers to the relatedness between two amino acid sequences. Those of ordinary skill in the art will appreciate that two sequences are generally considered to be “substantially identical” if they contain identical residues in corresponding positions. Amino acid sequences may be compared using any of a variety of algorithms well known in the art, including those available in commercial computer programs such as BLASTP, gapped BLAST, and PSI-BLAST, in the version in existence as of the date of filing. Exemplary programs are described in Altschul, et al., Basic local alignment search tool, *J. Mol. Biol.*, 215(3): 403-410, 1990; Altschul, et al., *Methods in Enzymology*; Altschul et al., *Nucleic Acids Res.* 25:3389-3402, 1997; Baxevanis et al., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Wiley, 1998; and Misener, et al., (eds.), *Bioinformatics Methods and Protocols* (Methods in Molecular Biology, Vol. 132), Humana Press, 1999. In some embodiments, the sequence identity between two amino acid sequences may be determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) as implemented in the Needle program of the

EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al., 2000, *Trends Genet.* 16: 276-277), in the version that exists as of the date of filing. The parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labeled “longest identity” (obtained using the `-nobrief` option) is used as the percent identity and is calculated as follows: $(\text{Identical Residues} \times 100) / (\text{Length of Alignment} - \text{Total Number of Gaps in Alignment})$

[85] In other embodiments, sequence identity may be determined using the Smith-Waterman algorithm, in the version that exists as of the date of filing.

[86] As used herein, “substantial homology” refers to a comparison between amino acid sequences. As will be appreciated by those of ordinary skill in the art, two sequences are generally considered to be “substantially homologous” if they contain homologous residues in corresponding positions. Homologous residues may be identical residues. Alternatively, homologous residues may be non-identical residues with appropriately similar structural and/or functional characteristics. For example, as is well known by those of ordinary skill in the art, certain amino acids are typically classified as “hydrophobic” or “hydrophilic” amino acids, and/or as having “polar” or “non-polar” side chains, and substitution of one amino acid for another of the same type may often be considered a “homologous” substitution.

[87] As is well known in this art, amino acid sequences may be compared using any of a variety of algorithms, including those available in commercial computer programs such as BLASTP, gapped BLAST, and PSI-BLAST, in existence as of the date of filing. Exemplary such programs are described in Altschul, et al., Basic local alignment search tool, *J. Mol. Biol.*, 215(3): 403-410, 1990; Altschul, et al., *Methods in Enzymology*; Altschul, et al., “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”, *Nucleic Acids Res.* 25:3389-3402, 1997; Baxevanis, et al., *Bioinformatics A Practical Guide to the Analysis of Genes and Proteins*, Wiley, 1998; and Misener, et al., (eds.), *Bioinformatics Methods and Protocols* (Methods in Molecular Biology, Vol. 132), Humana Press, 1999. In addition to identifying homologous sequences, the programs mentioned above typically provide an indication of the degree of homology.

[88] “Alkyl” or “alkyl group” refers to a fully saturated, straight or branched hydrocarbon chain having from one to twelve carbon atoms, and which is attached to the rest of the molecule by a single bond. Alkyls comprising any number of carbon atoms from 1 to 12 are included.

An alkyl comprising up to 12 carbon atoms is a C₁-C₁₂ alkyl, an alkyl comprising up to 10 carbon atoms is a C₁-C₁₀ alkyl, an alkyl comprising up to 6 carbon atoms is a C₁-C₆ alkyl and an alkyl comprising up to 5 carbon atoms is a C₁-C₅ alkyl. A C₁-C₅ alkyl includes C₅ alkyls, C₄ alkyls, C₃ alkyls, C₂ alkyls and C₁ alkyl (*i.e.*, methyl). A C₁-C₆ alkyl includes all moieties described above for C₁-C₅ alkyls but also includes C₆ alkyls. A C₁-C₁₀ alkyl includes all moieties described above for C₁-C₅ alkyls and C₁-C₆ alkyls, but also includes C₇, C₈, C₉ and C₁₀ alkyls. Similarly, a C₁-C₁₂ alkyl includes all the foregoing moieties, but also includes C₁₁ and C₁₂ alkyls. Non-limiting examples of C₁-C₁₂ alkyl include methyl, ethyl, *n*-propyl, *i*-propyl, *sec*-propyl, *n*-butyl, *i*-butyl, *sec*-butyl, *t*-butyl, *n*-pentyl, *t*-amyl, *n*-hexyl, *n*-heptyl, *n*-octyl, *n*-nonyl, *n*-decyl, *n*-undecyl, and *n*-dodecyl. Unless stated otherwise specifically in the specification, an alkyl group can be optionally substituted.

[89] “Alkylene” or “alkylene chain” refers to a fully saturated, straight or branched divalent hydrocarbon chain radical, having from one to forty carbon atoms. Non-limiting examples of C₂-C₄₀ alkylene include ethylene, propylene, *n*-butylene, ethenylene, propenylene, *n*-butenylene, propynylene, *n*-butynylene, and the like. The alkylene chain is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, an alkylene chain can be optionally substituted as described herein.

[90] “Alkenylene” or “alkenylene chain” refers to a straight or branched divalent hydrocarbon chain radical, having from two to forty carbon atoms, and having one or more carbon-carbon double bonds. Non-limiting examples of C₂-C₄₀ alkenylene include ethene, propene, butene, and the like. The alkenylene chain is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, an alkenylene chain can be optionally substituted.

[91] “Alkynyl” or “alkynyl group” refers to a straight or branched hydrocarbon chain having from two to twelve carbon atoms, and having one or more carbon-carbon triple bonds. Each alkynyl group is attached to the rest of the molecule by a single bond. Alkynyl group comprising any number of carbon atoms from 2 to 12 are included. An alkynyl group comprising up to 12 carbon atoms is a C₂-C₁₂ alkynyl, an alkynyl comprising up to 10 carbon atoms is a C₂-C₁₀ alkynyl, an alkynyl group comprising up to 6 carbon atoms is a C₂-C₆ alkynyl and an alkynyl comprising up to 5 carbon atoms is a C₂-C₅ alkynyl. A C₂-C₅ alkynyl includes C₅ alkynyls, C₄ alkynyls, C₃ alkynyls, and C₂ alkynyls. A C₂-C₆ alkynyl includes all moieties

described above for C₂-C₅ alkynyls but also includes C₆ alkynyls. A C₂-C₁₀ alkynyl includes all moieties described above for C₂-C₅ alkynyls and C₂-C₆ alkynyls, but also includes C₇, C₈, C₉ and C₁₀ alkynyls. Similarly, a C₂-C₁₂ alkynyl includes all the foregoing moieties, but also includes C₁₁ and C₁₂ alkynyls. Non-limiting examples of C₂-C₁₂ alkenyl include ethynyl, propynyl, butynyl, pentynyl and the like. Unless stated otherwise specifically in the specification, an alkyl group can be optionally substituted.

[92] “Alkynylene” or “alkynylene chain” refers to a straight or branched divalent hydrocarbon chain, having from two to forty carbon atoms, and having one or more carbon-carbon triple bonds. Non-limiting examples of C₂-C₄₀ alkynylene include ethynylene, propargylene and the like. The alkynylene chain is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, an alkynylene chain can be optionally substituted.

[93] “Carbocyclyl,” “carbocyclic ring” or “carbocycle” refers to a rings structure, wherein the atoms which form the ring are each carbon, and which is attached to the rest of the molecule by a single bond. Carbocyclic rings can comprise from 3 to 20 carbon atoms in the ring. Unless stated otherwise specifically in the specification, the carbocyclyl can be a monocyclic, bicyclic, tricyclic or tetracyclic ring system, which can include fused or bridged ring systems. Carbocyclic rings include aryls and cycloalkyl, cycloalkenyl, and cycloalkynyl as defined herein. Unless stated otherwise specifically in the specification, a carbocyclyl group can be optionally substituted. In some embodiments, the carbocyclyl divalent, and is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, a heterocyclyl group can be optionally substituted.

[94] “Cycloalkyl” refers to a stable non-aromatic monocyclic or polycyclic fully saturated hydrocarbon having from 3 to 40 carbon atoms and at least one ring, wherein the ring consists solely of carbon and hydrogen atoms, which can include fused or bridged ring systems. Monocyclic cycloalkyls include, for example, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl, and cyclooctyl. Polycyclic cycloalkyls include, for example, adamantyl, norbornyl, decalinyl, 7,7-dimethyl-bicyclo[2.2.1]heptanyl, and the like. In some embodiments, the cycloalkyl divalent and is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless otherwise stated specifically in the specification, a cycloalkyl group can be optionally substituted.

[95] “Cycloalkenyl” refers to a stable non-aromatic monocyclic or polycyclic hydrocarbon having from 3 to 40 carbon atoms, at least one ring having, and one or more carbon-carbon double bonds, wherein the ring consists solely of carbon and hydrogen atoms, which can include fused or bridged ring systems. Monocyclic cycloalkenyls include, for example, cyclopentenyl, cyclohexenyl, cycloheptenyl, cyclooctenyl, and the like. Polycyclic cycloalkenyl radicals include, for example, bicyclo[2.2.1]hept-2-enyl and the like. In some embodiments, cycloalkenyl is divalent and is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless otherwise stated specifically in the specification, a cycloalkenyl group can be optionally substituted.

[96] “Cycloalkynyl” refers to a stable non-aromatic monocyclic or polycyclic hydrocarbon having from 3 to 40 carbon atoms, at least one ring having, and one or more carbon-carbon triple bonds, wherein the ring consists solely of carbon and hydrogen atoms, which can include fused or bridged ring systems. Monocyclic cycloalkynyls include, for example, cycloheptynyl, cyclooctynyl, and the like. The cycloalkynyl is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless otherwise stated specifically in the specification, a cycloalkynyl group can be optionally substituted.

[97] “Aryl” refers to a hydrocarbon ring system comprising hydrogen, 6 to 40 carbon atoms and at least one aromatic ring. For purposes of this disclosure, the aryl can be a monocyclic, bicyclic, tricyclic or tetracyclic ring system, which can include fused or bridged ring systems. Aryls include, but are not limited to, aryl divalent radicals derived from aceanthrylene, acenaphthylene, acephenanthrylene, anthracene, azulene, benzene, chrysene, fluoranthene, fluorene, *as*-indacene, *s*-indacene, indane, indene, naphthalene, phenalene, phenanthrene, pleiadene, pyrene, and triphenylene. In some embodiments, the aryl divalent and is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, an aryl group can be optionally substituted.

[98] “Heterocyclyl,” “heterocyclic ring” or “heterocycle” refers to a stable 3- to 22-membered ring system which consists of two to fourteen carbon atoms and from one to eight heteroatoms selected from the group consisting of nitrogen, oxygen and sulfur. Heterocyclyl or heterocyclic rings include heteroaryls as defined below. Unless stated otherwise specifically in the specification, the heterocyclyl can be a monocyclic, bicyclic, tricyclic or tetracyclic ring system, which can include fused or bridged ring systems; and the

nitrogen, carbon or sulfur atoms in the heterocyclyl can be optionally oxidized; the nitrogen atom can be optionally quaternized; and the heterocyclyl can be partially or fully saturated. Examples of such heterocyclyl radicals include, but are not limited to, dioxolanyl, thienyl[1,3]dithianyl, decahydroisoquinolyl, imidazolanyl, imidazolidinyl, isothiazolidinyl, isoxazolidinyl, morpholanyl, octahydroindolyl, octahydroisoindolyl, 2-oxopiperazinyl, 2-oxopiperidinyl, 2-oxopyrrolidinyl, oxazolidinyl, piperidinyl, piperazinyl, 4-piperidonyl, pyrrolidinyl, succinimidyl, pyrazolidinyl, quinuclidinyl, thiazolidinyl, tetrahydrofuryl, trithianyl, tetrahydropyranyl, thiomorpholanyl, thiamorpholanyl, 1-oxo-thiomorpholanyl, and 1,1-dioxo-thiomorpholanyl. In some embodiments, the heterocyclyl is divalent and is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, a heterocyclyl group can be optionally substituted.

[99] “Heteroaryl” refers to a 5- to 22-membered aromatic ring comprising hydrogen atoms, one to fourteen carbon atoms, one to eight heteroatoms selected from the group consisting of nitrogen, oxygen and sulfur, and at least one aromatic ring. For purposes of this disclosure, the heteroaryl can be a monocyclic, bicyclic, tricyclic or tetracyclic ring system, which can include fused or bridged ring systems; and the nitrogen, carbon or sulfur atoms in the heteroaryl can be optionally oxidized; the nitrogen atom can be optionally quaternized. Examples include, but are not limited to, azepinyl, acridinyl, benzimidazolyl, benzothiazolyl, benzindolyl, benzodioxolyl, benzofuranyl, benzooxazolyl, benzothiazolyl, benzothiadiazolyl, benzo[*b*][1,4]dioxepinyl, 1,4-benzodioxanyl, benzonaphthofuranyl, benzoxazolyl, benzodioxolyl, benzodioxinyl, benzopyranyl, benzopyranonyl, benzofuranyl, benzofuranonyl, benzothienyl (benzothiophenyl), benzotriazolyl, benzo[4,6]imidazo[1,2-*a*]pyridinyl, carbazolyl, cinnolinyl, dibenzofuranyl, dibenzothiophenyl, furanyl, furanonyl, isothiazolyl, imidazolyl, indazolyl, indolyl, indazolyl, isoindolyl, indolinyl, isoindolinyl, isoquinolyl, indoliziny, isoxazolyl, naphthyridinyl, oxadiazolyl, 2-oxoazepinyl, oxazolyl, oxiranyl, 1-oxidopyridinyl, 1-oxidopyrimidinyl, 1-oxidopyrazinyl, 1-oxidopyridazinyl, 1-phenyl-1*H*-pyrrolyl, phenazinyl, phenothiazinyl, phenoxazinyl, phthalazinyl, pteridinyl, purinyl, pyrrolyl, pyrazolyl, pyridinyl, pyrazinyl, pyrimidinyl, pyridazinyl, quinazoliny, quinoxaliny, quinoliny, quinuclidinyl, isoquinoliny, tetrahydroquinoliny, thiazolyl, thiadiazolyl, triazolyl, tetrazolyl, triazinyl, and thiophenyl (i.e. thienyl). In some embodiments, the heteroaryl is divalent and is attached, directly or indirectly, to the CPP through a single

bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, a heteroaryl group can be optionally substituted.

[100] The term “ether” used herein refers to a divalent moiety having a formula $-(R_1)_m-O-(R_2)_n-$ wherein each of m, n, and z are independently selected from 1 to 40, and R1 and R2 are independently selected from an alkylene. Examples include polyethylene glycol. The ether is attached, directly or indirectly, to the CPP through a single bond and, directly or indirectly, to the TP through a single bond. Unless stated otherwise specifically in the specification, the ether can be optionally substituted.

[101] The term “substituted” used herein means any of the above groups (*i.e.*, alkylene, alkenylene, alkynylene, aryl, carbocyclyl, cycloalkyl, cycloalkenyl, cycloalkynyl, heterocyclyl, heteroaryl, and/or ether) wherein at least one hydrogen atom is replaced by a bond to a non-hydrogen atoms such as, but not limited to: a deuterium atom; a halogen atom such as F, Cl, Br, and I; an oxygen atom in groups such as hydroxyl groups, alkoxy groups, and ester groups; a sulfur atom in groups such as thiol groups, thioalkyl groups, sulfone groups, sulfonyl groups, and sulfoxide groups; a nitrogen atom in groups such as amines, amides, alkylamines, dialkylamines, arylamines, alkylarylamines, diarylamines, N-oxides, imides, and enamines; a silicon atom in groups such as trialkylsilyl groups, dialkylarylsilyl groups, alkylarylsilyl groups, and triarylsilyl groups; and other heteroatoms in various other groups. “Substituted” also means any of the above groups in which one or more hydrogen atoms are replaced by a higher-order bond (e.g., a double- or triple-bond) to a heteroatom such as oxygen in oxo, carbonyl, carboxyl, and ester groups; and nitrogen in groups such as imines, oximes, hydrazones, and nitriles. For example, “substituted” includes any of the above groups in which one or more hydrogen atoms are replaced with $-NR_gR_h$, $-NR_gC(=O)R_h$, $-NR_gC(=O)NR_gR_h$, $-NR_gC(=O)OR_h$, $-NR_gSO_2R_h$, $-OC(=O)NR_gR_h$, $-OR_g$, $-SR_g$, $-SOR_g$, $-SO_2R_g$, $-OSO_2R_g$, $-SO_2OR_g$, $=NSO_2R_g$, and $-SO_2NR_gR_h$. “Substituted” also means any of the above groups in which one or more hydrogen atoms are replaced with $-C(=O)R_g$, $-C(=O)OR_g$, $-C(=O)NR_gR_h$, $-CH_2SO_2R_g$, $-CH_2SO_2NR_gR_h$. In the foregoing, R_g and R_h are the same or different and independently hydrogen, alkyl, alkenyl, alkynyl, alkoxy, alkylamino, thioalkyl, aryl, aralkyl, cycloalkyl, cycloalkenyl, cycloalkynyl, cycloalkylalkyl, haloalkyl, haloalkenyl, haloalkynyl, heterocyclyl, *N*-heterocyclyl, heterocyclylalkyl, heteroaryl, *N*-heteroaryl and/or heteroarylalkyl. “Substituted” further means any of the above groups in which one or more hydrogen atoms are replaced by a bond to an amino, cyano, hydroxyl, imino, nitro, oxo, thioxo, halo, alkyl, alkenyl, alkynyl, alkoxy, alkylamino, thioalkyl,

aryl, aralkyl, cycloalkyl, cycloalkenyl, cycloalkynyl, cycloalkylalkyl, haloalkyl, haloalkenyl, haloalkynyl, heterocyclyl, *N*-heterocyclyl, heterocyclylalkyl, heteroaryl, *N*-heteroaryl and/or heteroarylalkyl group. In addition, each of the foregoing substituents can also be optionally substituted with one or more of the above substituents. Further, those skilled in the art will recognize that “substituted” also encompasses instances in which one or more hydrogen atoms on any of the above groups are replaced by a substituent listed in this paragraph, and the substituent then forms a covalent bond with the CPP or TP. The resulting bonding group can be considered a “substituent.” For example, in certain embodiments, any of the above groups can be substituted at a first position with a carboxylic acid (i.e., -C(=O)OH) which forms an amide bond with an appropriate amino acid CPP (e.g., lysine), and also substituted at a second position with either an electrophilic group (e.g., -C(=O)H, -CO₂R_g, -halide, etc.) which forms a bond with the N-terminus of TP or alternatively a nucleophilic group (-NH₂, -NHR_g, -OH, etc.) which forms a bond with the C-terminus of TP. The resulting bond, e.g., amide bond, can be considered a “substituent.” In some embodiments, the second position is substituted with a thiol group which forms a disulfide bond with a cysteine (or amino acid analog having a thiol group) in TP. The resulting disulfide is encompassed by the term substituent.

[102] As used herein, the symbol “” (hereinafter can be referred to as “a point of attachment bond”) denotes a bond that is a point of attachment between two chemical entities, one of which is depicted as being attached to the point of attachment bond and the other of which is not depicted as being attached to the point of attachment bond. For example, “XY” indicates that the chemical entity “XY” is bonded to another chemical entity via the point of attachment bond. Furthermore, the specific point of attachment to the non-depicted chemical entity can be specified by inference. For example, the compound CH₃-R³, wherein R³ is H or “XY” infers that when R³ is “XY”, the point of attachment bond is the same bond as the bond by which R³ is depicted as being bonded to CH₃.

Compounds

[103] Disclosed herein, in various embodiments, are compounds for treating Mitochondrial Neurogastrointestinal Encephalopathy Syndrome (MNGIE). The compounds are designed to deliver a moiety with thymidine phosphorylate activity intracellularly to MNGIE patients or patients with Mutations in the *TYMP*. By doing so, the compounds reduce the toxic levels of

nucleosides that would otherwise accumulate in such patients. In some embodiments, the present compounds reduce toxic nucleoside levels by about 1%, about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 99%, or about 100%, inclusive of all values and ranges therebetween.

[104] In various embodiments, the compounds disclosed herein have a thymidine phosphorylase activity and cell penetrating activity, such that the compounds are able to traverse the cell membrane and reduce thymidine levels *in vivo*. In some embodiments, the compounds comprise: a) at least one cell-penetrating peptide (CPP) moiety; and b) at least one thymidine phosphorylase, or an active fragment or analog thereof (TP), wherein the CPP is coupled, directly or indirectly, to TP. In some embodiments, the compounds comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more TP moieties. In some embodiments, the compounds comprise one TP moiety. In some embodiments, the compounds comprise two TP moieties. As used herein, “coupled” can refer to a covalent or non-covalent association between the CPP to the TP, including fusion of the CPP to the TP and chemical conjugation of the CPP to the TP. A non-limiting example of a means to non-covalently attach the CPP to the TPP is through the streptavidin/biotin interaction, e.g., by conjugating biotin to CPP and fusing TP to streptavidin. In the resulting compound, the CPP is coupled to the TP via non-covalent association between biotin and streptavidin.

[105] In some embodiments, the CPP is conjugated, directly or indirectly, to the TP to thereby form a CPP-TP conjugate. Conjugation of the TP to the CPP may occur at any appropriate site on these moieties. For example, in some embodiments, the N-terminus or C-terminus of the TP may be conjugated to the C-terminus, the N-terminus, or a side chain of an amino acid in the CPP. In some embodiments, the CPP may be conjugated to the side change of an amino acid in TP.

[106] In some embodiments, the TP is fused to the CPP. Fusion proteins, as used herein, refer to constructs where a linear CPP moiety is fused to the N- and/or C-terminus of the TP moiety. Such fusion protein may alternatively be described as having a cell penetrating domain and a thymidine domain. Methods of fusing polypeptides are well-known in the art. Such fusion constructs may be prepared by recombinant techniques. A recombinantly-produced TP-CPP fusion protein, in accordance with certain embodiments of the disclosure, includes the TP component and the linear CPP component associated with one another by genetic fusion. For example, the fusion protein may be generated by translation of a polynucleotide encoding the

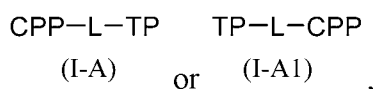
TP cloned in-frame with the linear CPP component (or vice versa). Such a fusion protein may contain one or more copies of CPP attached to the N-terminus and/or the C-terminus of the TP component. In some embodiments, a CPP component is independently attached to both the N- and C-terminus of the TP component.

[107] In other embodiments, the TP may be chemically conjugated to the CPP through a side chain of an amino acid on TP. In still other embodiments, the TP may be conjugated to the CPP through a side chain of an amino acid on the CPP. Any amino acid side chain on the CPP and/or TP which is capable of forming a covalent bond, or which may be so modified, can be used to link TP to the CPP. The amino acid on the CPP can be a natural or non-natural amino acid. In some embodiments, the amino acid on the CPP used to conjugate the TP is aspartic acid, glutamic acid, glutamine, asparagine, lysine, ornithine, 2,3-diaminopropionic acid, or analogs thereof, wherein the side chain is substituted with a bond to the TPP or linker. In particular embodiments, the amino is lysine, or an analog thereof. In other embodiments, the amino acid is glutamic acid, or an analog thereof. In further embodiments, the amino acid is aspartic acid, or an analog thereof.

[108] In some embodiments of the present disclosure, the compounds further comprise a linker (L), which conjugates the CPP to TP. In some embodiments, L conjugates the CPP to the N-terminus or the C-terminus of the TP. In a certain embodiment, L conjugates the CPP to the N-terminus of the TP.

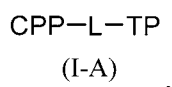
[109] In some embodiments, the CPP is conjugated to the TP through a side chain of an amino acid on the TP. Any appropriate side chain of an amino acid of TP which is capable of forming a covalent bond with the CPP, or which may be so modified, can be used to conjugate the CPP to TP. The amino acid may be a constituent of native TP or a non-native amino acid. That is, in some embodiments, TP can include a non-native amino acid which provides a handle to conjugate the CPP. In particular embodiments, the amino acid is glutamine, asparagine, lysine, cysteine, tryptophan, or analogs thereof.

[110] In some embodiments, the CPP is cyclic (as described herein), and referred to herein as a cCPP. There are numerous possible configurations for the compounds disclosed herein. In certain embodiments, the compounds of the disclosure are exocyclic compounds wherein TP is conjugated to the side chain of an amino acid in the cCPP. In some embodiments, the compounds disclosed herein have structure (i.e., exocyclic) according to Formula I-A or Formula I-A1:



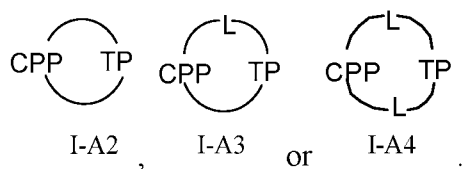
wherein L is a covalently bound to the side chain of an amino acid on the CPP and to the N-terminus of the TP, an amino acid side chain of TP, or the C-terminus of the TP.

[111] In certain embodiments, the compounds (e.g., exocyclic compounds) disclosed herein have a structure according to Formula I-A:

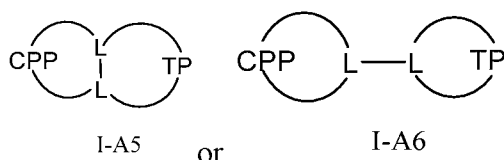


wherein L is a covalently bound to the side chain of an amino acid on the CPP and to the N-terminus of the TP.

[112] In some embodiments of the present disclosure, the CPP and TP together are cyclic (referred to herein as an “endocyclic compound”). In various non-limiting embodiments, the endocyclic compounds disclosed herein have a structure according to Formula I-A2, Formula I-A3, or Formula I-A4:



[113] In other embodiments, the TP moiety is cyclic and the CPP is a cyclic, and together they form a fused bicyclic system (referred to herein as a “bicyclic compound”). In various non-limiting embodiments, the endocyclic compounds disclosed herein have a structure according to Formula I-A5 and I-A6:



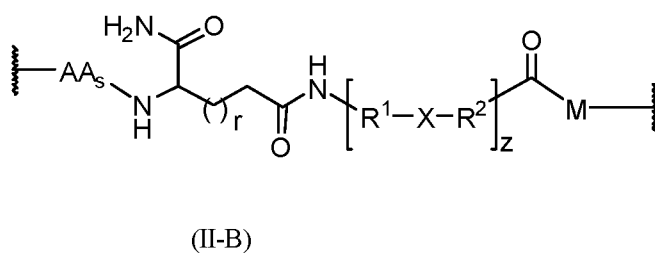
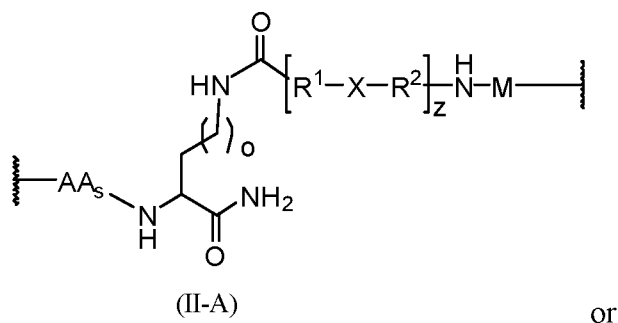
[114] L may be any appropriate moiety which conjugates CPP (e.g., as described herein) to a TP moiety. Thus, prior to conjugation to the CPP and TP, the linker has two or more functional groups, each of which are independently capable of forming a covalent bond to the CPP moiety and the TP moiety. In various embodiments of the present disclosure, L is covalently bound to the N-terminus of TP or the C-terminus of TP. In some embodiments, L is covalently bound to the N-terminus of TP. In other embodiments, L is covalently bound to the C-terminus of TP. In still other embodiments, L is covalently bound to the side chain of an amino acid in TP.

[115] In various embodiments of the present disclosure, L comprises (i) one or more D or L amino acids, each of which is optionally substituted; (ii) alkylene, alkenylene, alkynylene, carbocyclyl, or heterocyclyl, each of which is optionally substituted; or (iii) $-(R^1-X-R^2)_z-$, wherein each of R^1 and R^2 , at each instance, are independently selected from alkylene, alkenylene, alkynylene, carbocyclyl, and heterocyclyl, each X is independently NR^3 , $-NR^3C(O)-$, S, and O, wherein R^3 is H, alkyl, alkenyl, alkynyl, carbocyclyl, or heterocyclyl, each of which is optionally substituted, and z is an integer from 1 to 50; or (iv) combinations thereof. In some embodiments, L comprises one or more D or L amino acids, each of which is optionally substituted. In other embodiments, L comprises alkylene, alkenylene, alkynylene, carbocyclyl, or heterocyclyl, each of which is optionally substituted. In still other embodiments, L comprises $-(R^1-X-R^2)_z-$, wherein each of R^1 and R^2 , at each instance, are independently selected from alkylene, alkenylene, alkynylene, carbocyclyl, and heterocyclyl, each X is independently NR^3 , $-NR^3C(O)-$, S, and O, wherein R^3 is H, alkyl, alkenyl, alkynyl, carbocyclyl, or heterocyclyl, each of which is optionally substituted, and z is an integer from 1 to 50; or combinations thereof. In certain embodiments, L is an ether, which is optionally substituted. In more specific embodiments, L comprises $-(CH_2-O-CH_2)_z-$, wherein Z is an integer from 1-50. In more specific embodiments, L comprises $-(CH_2-O-CH_2)_z-$, wherein Z is an integer from 1-25 (e.g., 12), and one or more D or L amino acids, such as and lysine. For example, in various embodiments, L comprises a polyethylene glycol moiety, having from 1 to 50 ethylene glycol units, and a lysine residue. In other specific embodiments, L comprises $-(CH_2-S-CH_2)_z-$, wherein Z is an integer from 1-50. In still other specific embodiments, L comprises $-(CH_2-NR^3-CH_2)_z-$, wherein R^3 is H, $-C(O)$, alkyl, alkenyl, alkynyl, carbocyclyl, or heterocyclyl, each of which is optionally substituted, and z is an integer from 1-50, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and 50, inclusive of all subranges therebetween. In some embodiments, z is an integer from 10-15. In a specific embodiment, z is 12.

[116] As discussed above, L or M may be covalently bound to TP at any suitable location on TP. In various embodiments of the present disclosure, L or M is covalently bound to the N-terminus of TP or the C-terminus of TP. In another embodiment, L or M is covalently bound to the N-terminus of TP. In some embodiments, L or M is covalently bound to an amino acid side chain of TP.

[117] In some embodiments, L is bound to the side chain of aspartic acid, glutamic acid, glutamine, asparagine, or lysine, or a modified side chain of glutamine or asparagine (e.g., a reduced side chain having an amino group), on the CPP or TP. In particular embodiments, the L is bound to the side chain of lysine on the CPP.

[118] In some embodiments, L has a structure according to Formula II-A or Formula II-B:

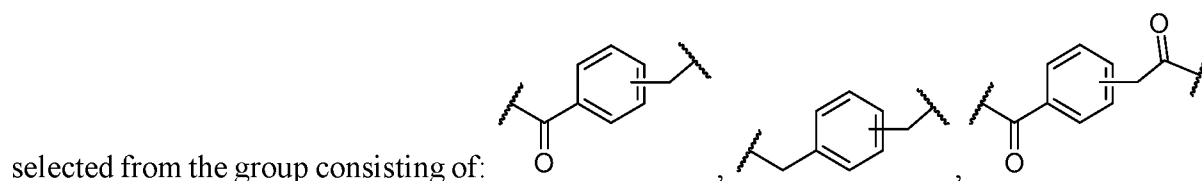


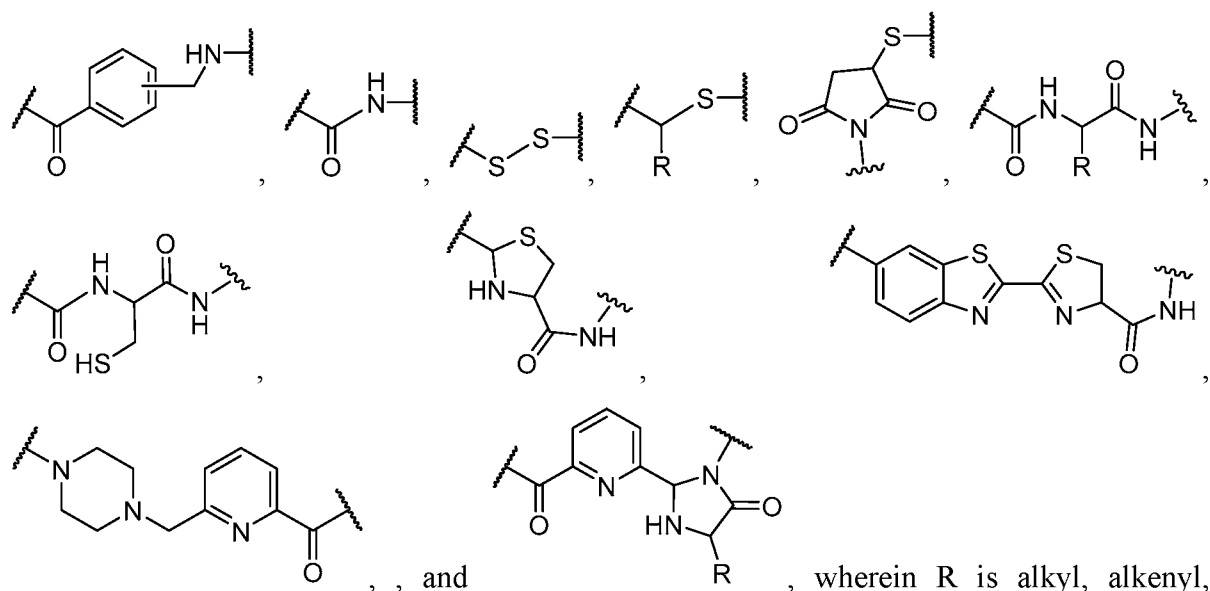
wherein each $-(R^1-X-R^2)_z-$ is defined as above.

[119] In some embodiments, each of R^1 and R^2 , at each instance, are independently selected from alkylene, alkenylene, alkynylene, carbocyclyl, and heterocyclyl, each of which is optionally substituted.

[120] In some embodiments, each X is independently NR^3 , $-NR^3C(O)-$, S, and O, and wherein R^3 is independently selected from H, alkyl, alkenyl, alkynyl, carbocyclyl, and heterocyclyl, each of which is optionally substituted.

[121] In some embodiments, M is absent or a group bound to an amino acid on TP. In various embodiments, M is present and comprises an alkylene, alkenylene, alkynylene, carbocyclyl, or heterocyclyl, each of which is optionally substituted. In some embodiments, M is present and





[122] In some embodiments, AA_s is a side chain or terminus of an amino acid on the CPP. Non-limiting examples of AA_s include aspartic acid, glutamic acid, glutamine, asparagine, or lysine, or a modified side chain of glutamine or asparagine (e.g., a reduced side chain having an amino group).

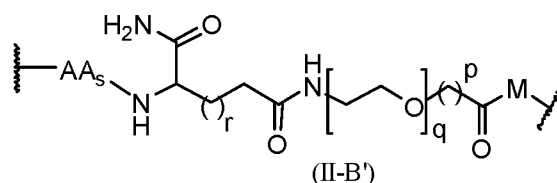
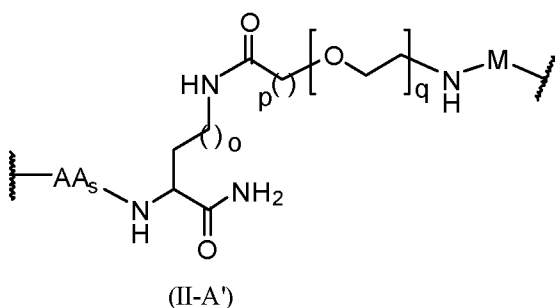
[123] In some embodiments, o is an integer from 0 to 10, e.g., 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, inclusive of all values and subranges therebetween. In other embodiments, o is 0, 1, 2, or 3.

[124] In some embodiments, u is 0 or 1. In some embodiments, u is 0. In other embodiments u is 1.

[125] In some embodiments p is 1 or 2. In some embodiments, p is 1. In other embodiments, p is 2.

[126] In some embodiments, r is 0 or 1. In some embodiments, r is 0. In other embodiments, r is 1.

[127] In some embodiments, L has a structure according to Formula II-A' or II-B':

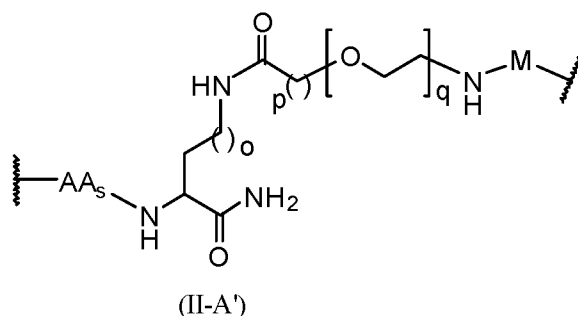


wherein each of M, AA_s, u, o, p, and r are defined above.

[128] In some embodiments, q is an integer from 1 to 50, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and 50, inclusive of all ranges and values therebetween. In other embodiments, q is an integer from 5-20. In other embodiments, q is an integer from 10-15. In a specific embodiment, q is 12.

[129] In certain embodiments, o is 0, 1, 2, or 3. In certain other embodiments, r is 0 or 1.

[130] In certain embodiments of the present disclosure, L is Formula II-A':



wherein

M is absent or a group bound to an amino acid on TP;

AA_s is a side chain or terminus of an amino acid on the CPP;

u is 0 or 1;

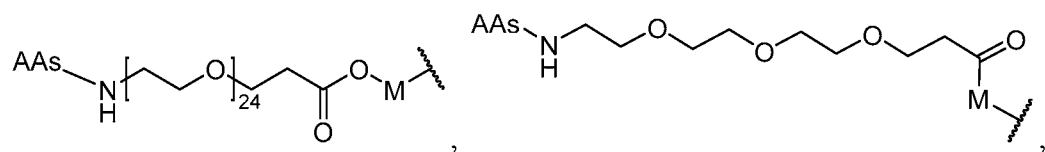
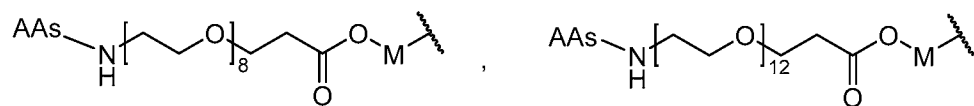
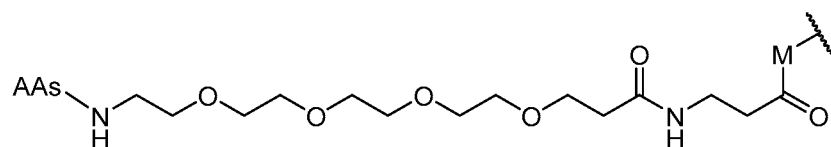
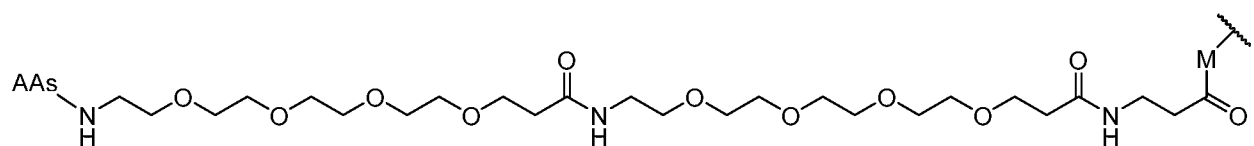
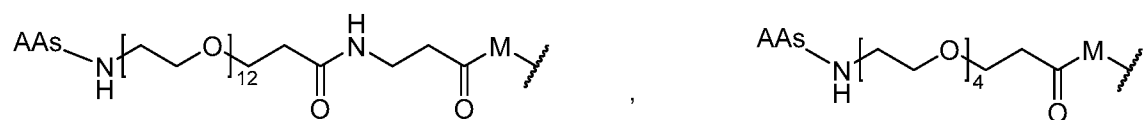
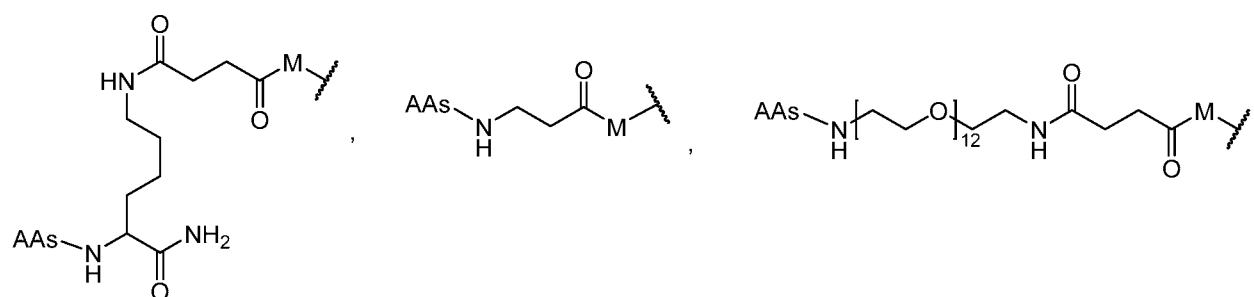
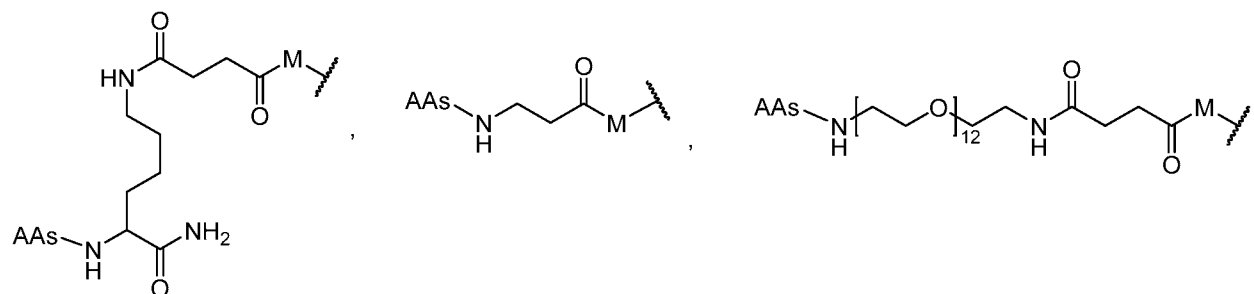
o is 0, 1, 2, or 3;

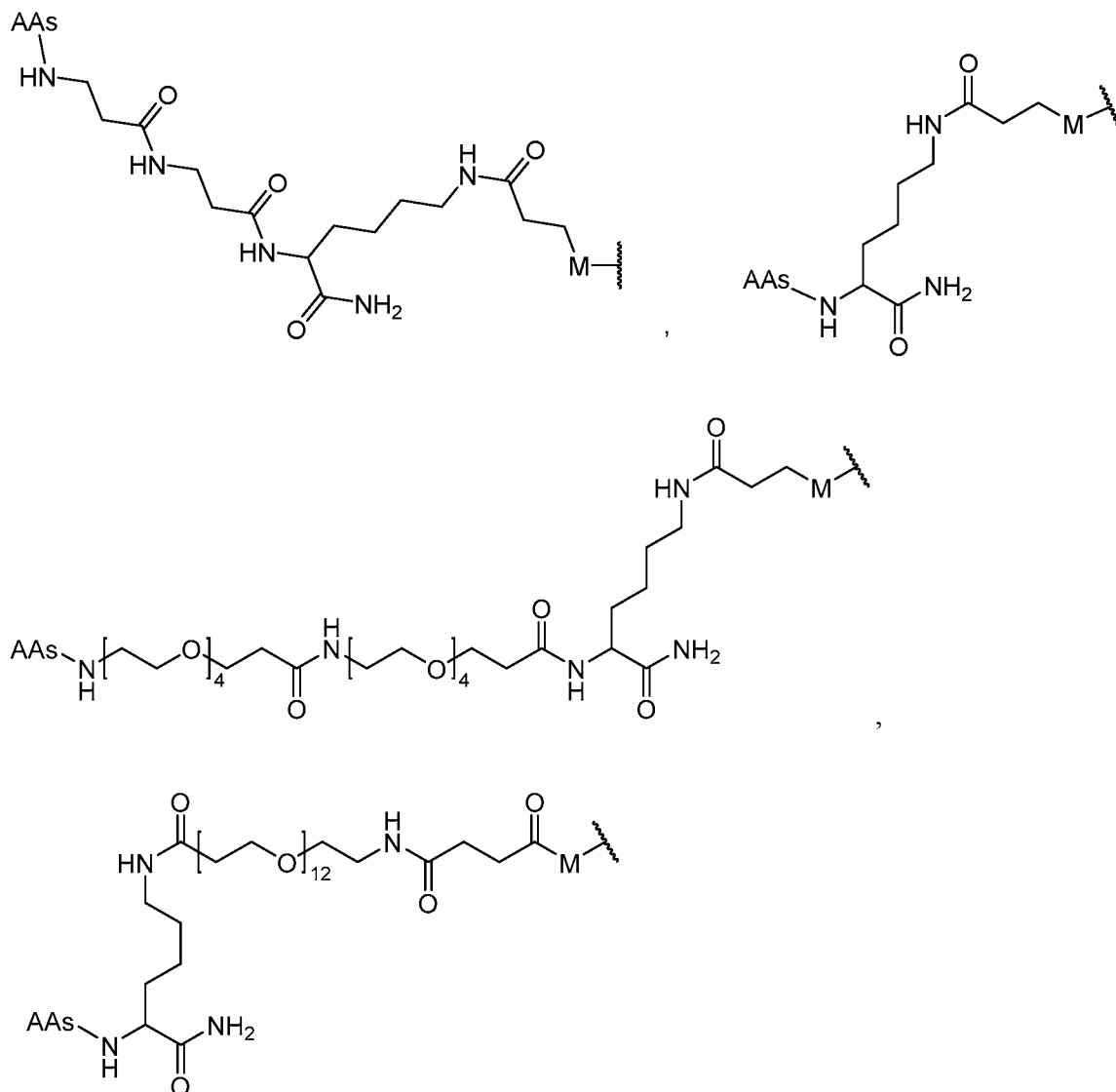
p is 1 or 2; and

q is an integer from 10 to 15.

[131] In some embodiments of the present disclosure, r is 0, p is 2, and q is 12. In other embodiments, r is 0. In still other embodiments, p is 2. In further embodiments, q is 12.

[132] Other non-limiting examples of suitable L groups include:

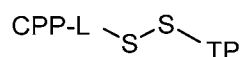




[133] In some embodiments, the L contains a group which may be cleaved after cytosolic uptake of the compounds of the disclosure to release TP. Non-limiting examples of physiologically cleavable linking group include carbonate, thiocarbonate, thioester, disulfide, sulfoxide, hydrazine, protease-cleavable dipeptide linker, and the like.

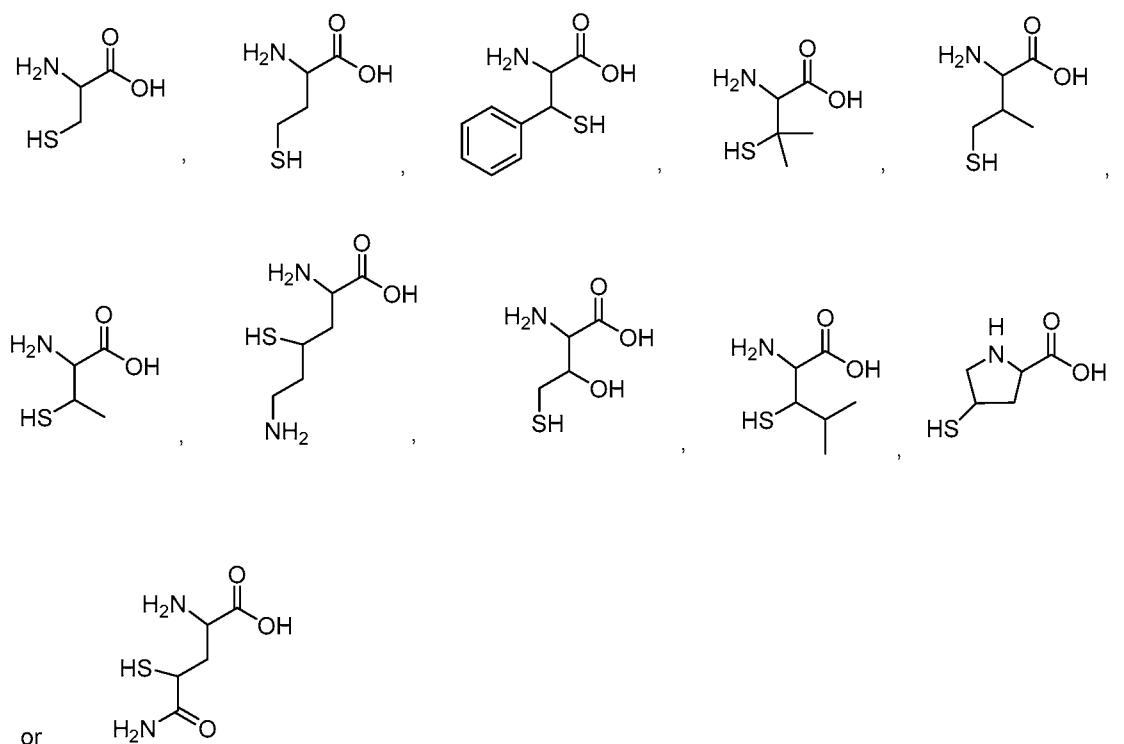
[134] In certain embodiments, a precursor to L also contains a thiol group, which forms a disulfide bond with the side chain of cysteine or cysteine analog located on TP.

[135] Accordingly, in various embodiments, the compounds disclosed herein (e.g., the compounds for Formula (I-A)) have the following structure:



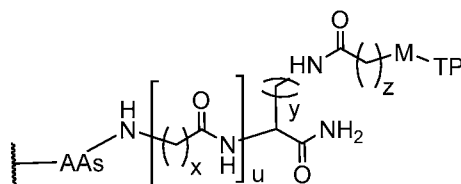
[136] In some embodiments, the disulfide bond is formed between a thiol group on L, and the side chain of cysteine or an amino acid analog having a thiol group on TP. Such thiol containing side chains may be located on native amino acids of wild-type TP, or such thiol

containing amino acids may be introduced to TP. Non-limiting examples of amino acid analogs having a thiol group which can be used with the polypeptide conjugates disclosed herein include:



[137] One skilled in the art will recognize that the amino acid analogs depicted above are shown as precursors, i.e., prior to incorporation into the compounds. When incorporated in the compounds of the present disclosure, the N- and C-termini are independently substituted to form peptide bonds, and the hydrogen on the thiol group is replaced with a bond to another sulfur atom to thereby form a disulfide.

[138] In some embodiments, L is Formula II-C':



II-C'

wherein:

AA_s is a side chain or terminus of an amino acid on the CPP;

M is defined above;

z is an integer from 0 to 10;

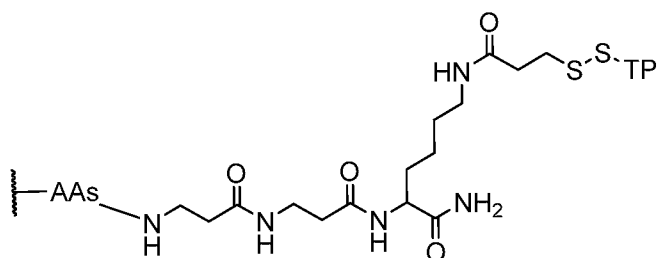
y is an integer from 0 to 10;

x is an integer from 0 to 10; and

u is an integer from 1 to 50.

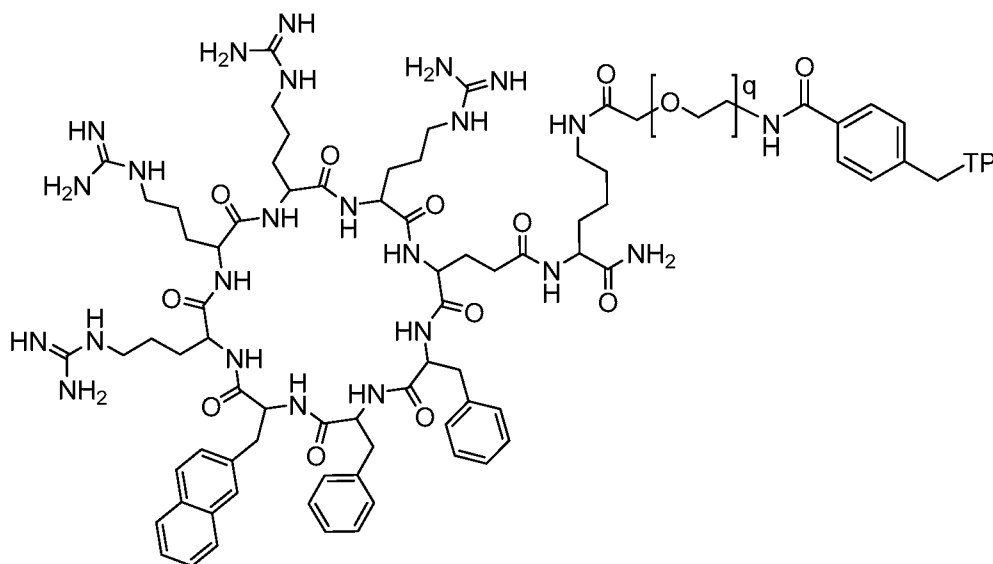
[139] In some embodiments, M is a physiologically cleavable bond. In some embodiments, M is disulfide.

[140] In some embodiments, L is

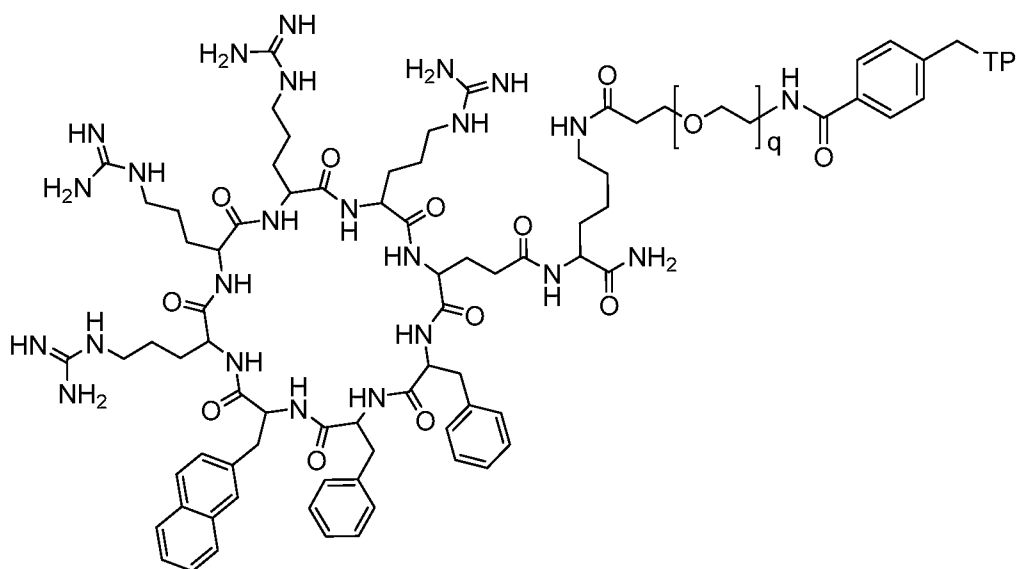


[141] In particular embodiments, a disulfide bond is formed between a thiol group on L, and the side chain of cysteine on TP. In some embodiments, said cysteine may be a constituent of wild type TP or TP may be modified to include cysteine or an amino acid analog having a thiol group. In other embodiments, any suitable functional group of TP may be modified to form a thiol group for bonding to L.

[142] In more specific embodiments of the present disclosure, the compound has a structure according to Formula V-A1, V-A2, or V-A3:

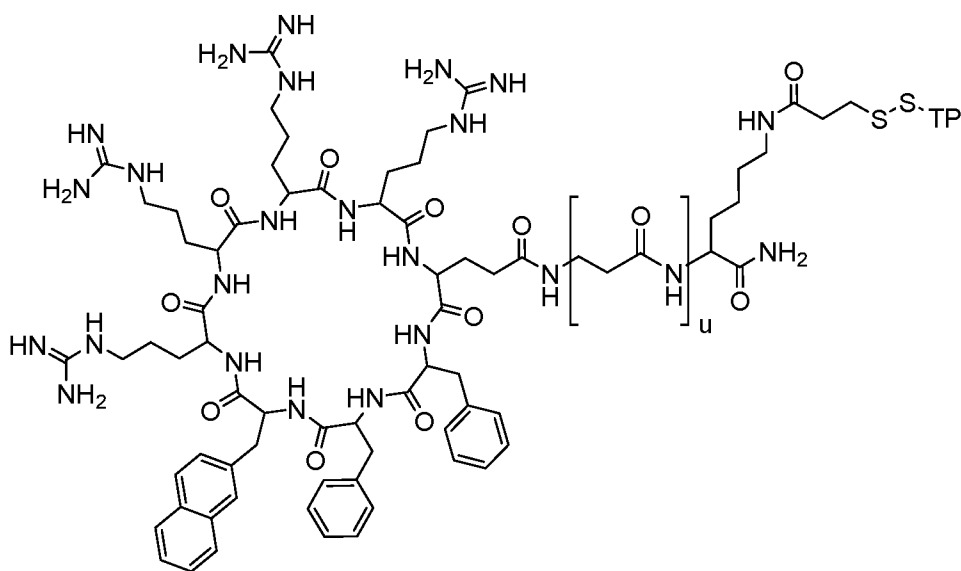


(V-A1)



(V-A2).

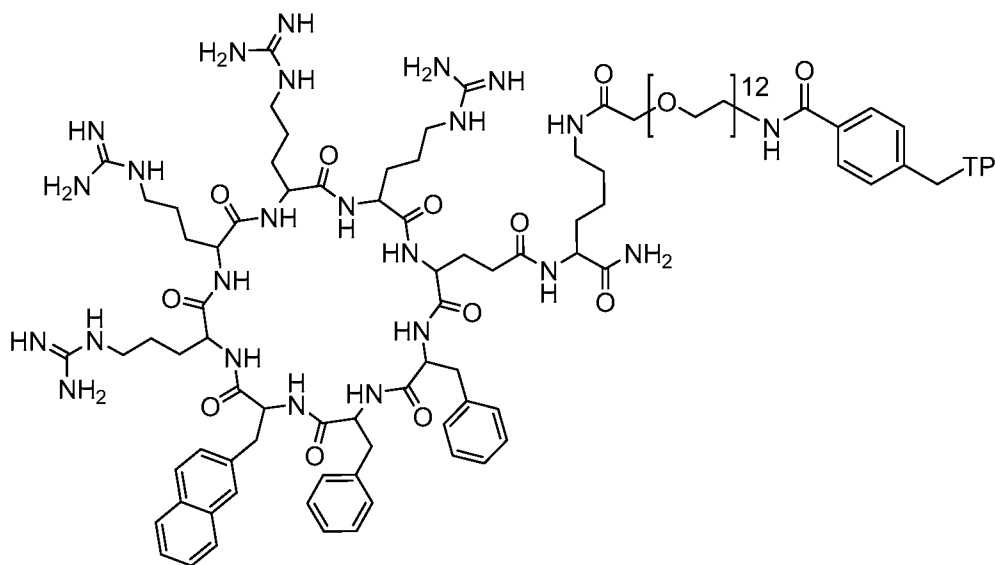
or



(V-A3)

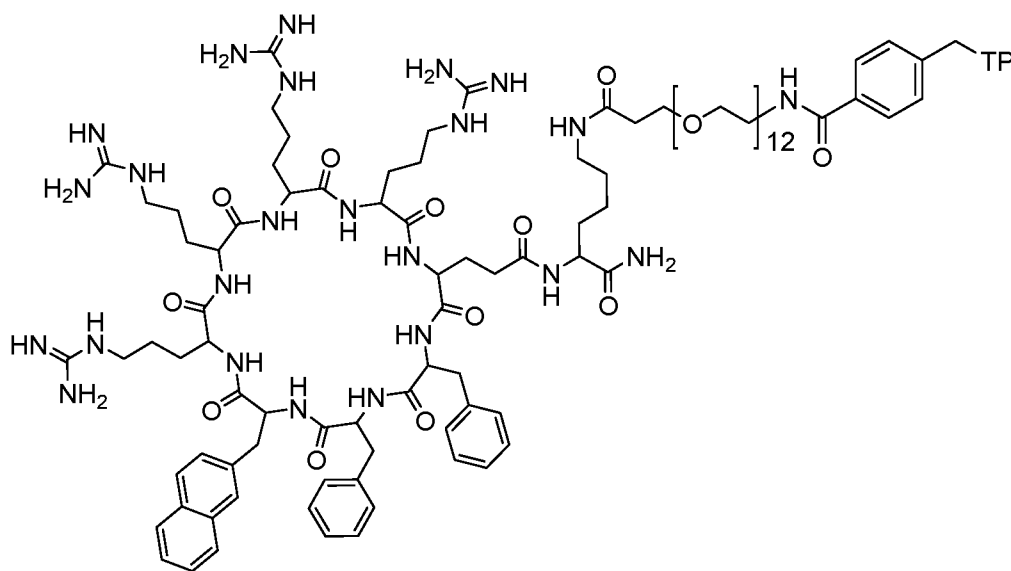
[143] In various embodiments, q may be any integer described above, e.g., an integer in the range of from 10 to 15.

[144] In specific embodiments the present disclosure, the compound has a structure according to Formula V-B1 or V-B2:



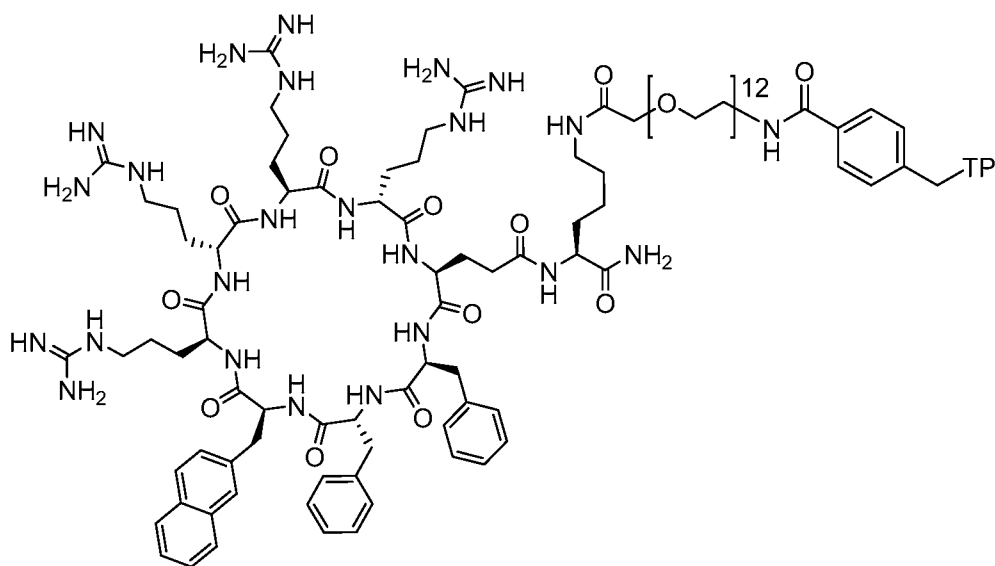
(V-B1)

or



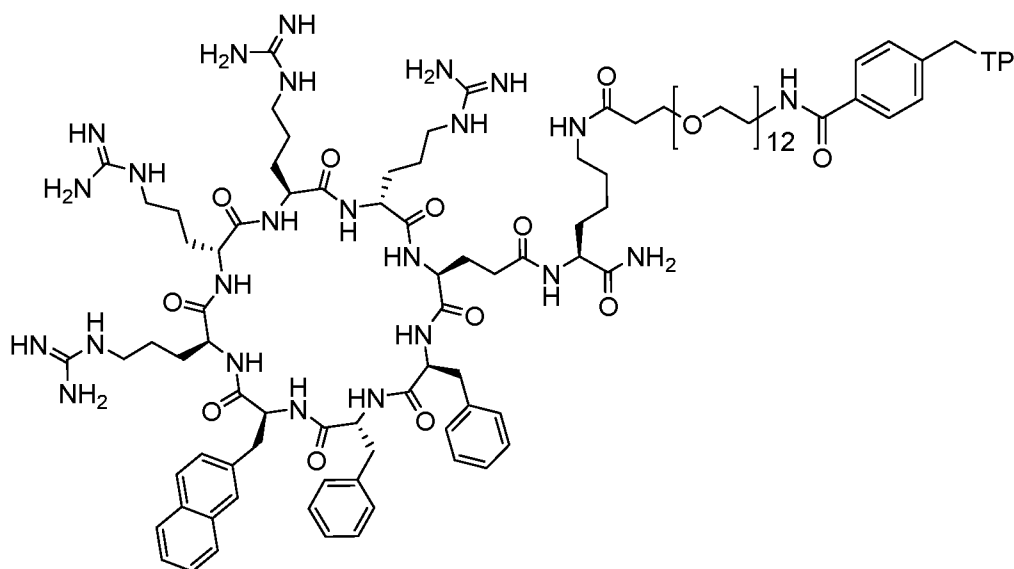
(V-B2)

[145] In other specific embodiments, the present disclosure provides a compound having the structure of Formula V-B3 or V-B4:

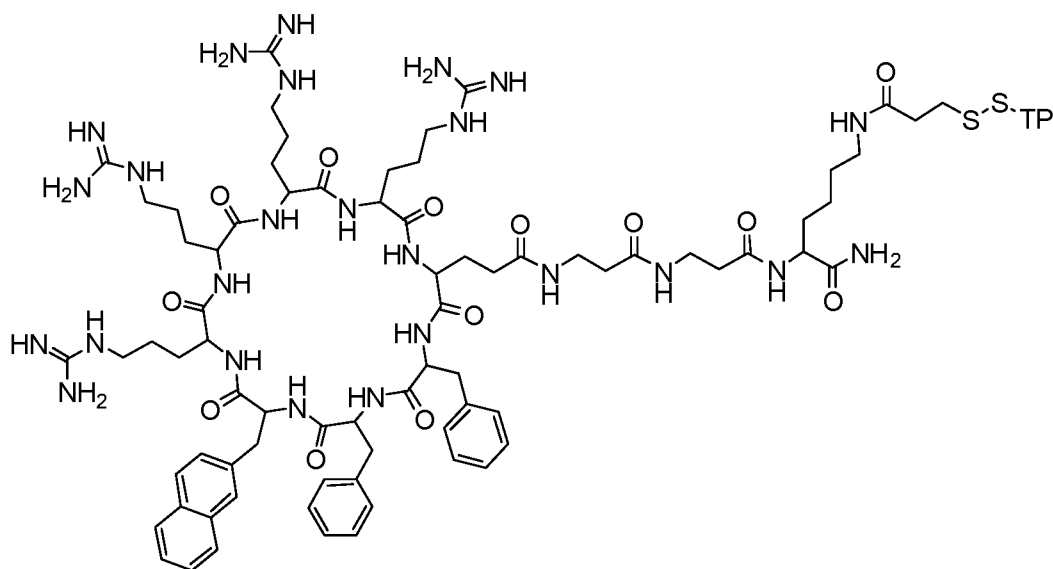


(V-B3)

or



(V-B4)



(V-B5)

[146] The TP in any of the above structures may be any TP disclosed herein, including SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13. Further, the TP may comprise a water-soluble polymer. In some embodiments, the water-soluble polymer comprises a PEG residue. In some embodiments, the PEG residue has a molecular weight ranging from about 1 kDa to about 100 kDa, e.g., from about 1 kDa to about 20 kDa, including about 10 kDa.

Cell-Penetrating Peptides

[147] As discussed above, the compounds disclosed herein comprise cell-penetrating peptides (CPPs).

[148] The CPP may be or include any amino sequence which facilitates cellular uptake of the compounds disclosed herein. Suitable CPPs for use in the compounds and methods described herein can include naturally occurring sequences, modified sequences, and synthetic sequences. In embodiments, the total number of amino acids in the CPP may be in the range of from 4 to about 20 amino acids, e.g., about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, and about 19 amino acids, inclusive of all ranges and subranges therebetween. In some embodiments, the CPPs disclosed herein comprise about 4 to about 13 amino acids. In particular embodiments, the CPPs disclosed herein comprise about 6 to about 10 amino acids, or about 6 to about 8 amino acids.

[149] Each amino acid in the CPP may be a natural or non-natural amino acid. The term “non-natural amino acid” refers to an organic compound that is a congener of a natural amino

acid in that it has a structure similar to a natural amino acid so that it mimics the structure and reactivity of a natural amino acid. The non-natural amino acid can be a modified amino acid, and/or amino acid analog, that is not one of the 20 common naturally occurring amino acids or the rare natural amino acids selenocysteine or pyrrolysine. Non-natural amino acids can also be the D-isomer of the natural amino acids. Examples of suitable amino acids include, but are not limited to, alanine, allosoleucine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, naphthylalanine, phenylalanine, proline, pyroglutamic acid, serine, threonine, tryptophan, tyrosine, valine, a derivative, or combinations thereof. These, and others, are listed in the Table 1 along with their abbreviations used herein.

Table 1. Amino Acid Abbreviations

Amino Acid	Abbreviations*	
	L-amino acid	D-amino acid
Alanine	Ala (A)	ala (a)
Allo-isoleucine	AlIle	aile
Arginine	Arg (R)	arg (r)
Asparagine	Asn (N)	asn (n)
aspartic acid	Asp (D)	asp (d)
Cysteine	Cys (C)	cys (c)
Cyclohexylalanine	Cha	cha
2,3-diaminopropionic acid	Dap	dap
4-fluorophenylalanine	Fpa (Σ)	pfa
glutamic acid	Glu (E)	glu (e)
glutamine	Gln (Q)	gln (q)
glycine	Gly (G)	gly (g)
histidine	His (H)	his (h)
Homoproline (aka pipecolic acid)	Pip (Θ)	pip (θ)
isoleucine	Ile (I)	ile (i)
leucine	Leu (L)	leu (l)
lysine	Lys (K)	lys (k)
methionine	Met (M)	met (m)
naphthylalanine	Nal (Φ)	nal (ϕ)

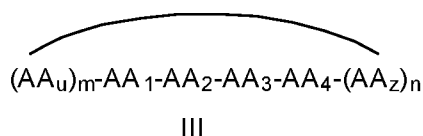
Amino Acid	Abbreviations*	
	L-amino acid	D-amino acid
norleucine	Nle (Ω)	nle
phenylalanine	Phe (F)	phe (F)
phenylglycine	Phg (Ψ)	phg
4-(phosphonodifluoromethyl)phenylalanine	F ₂ Pmp (Λ)	f ₂ pmp
proline	Pro (P)	pro (p)
sarcosine	Sar (Ξ)	sar
selenocysteine	Sec (U)	sec (u)
serine	Ser (S)	ser (s)
threonine	Thr (T)	thr (y)
tyrosine	Tyr (Y)	tyr (y)
tryptophan	Trp (W)	trp (w)
valine	Val (V)	val (v)
Tert-butyl-alanine	Tle	tle
Penicillamine	Pen	pen
Homoarginine	HomoArg	homoarg
Nicotinyl-lysine	Lys(NIC)	lys(NIC)
Trifluoroacetyl-lysine	Lys(TFA)	lys(TFA)
Methyl-leucine	MeLeu	meLeu
3-(3-benzothienyl)-alanine	Bta	bta

* single letter abbreviations: when shown in capital letters herein it indicates the L-amino acid form, when shown in lower case herein it indicates the D-amino acid form.

[150] Non-limiting examples of linear CPPs include Polyarginine (e.g., R₉ or R₁₁), Antennapedia sequences, HIV-TAT, Penetratin, Antp-3A (Antp mutant), Buforin II, Transportan, MAP (model amphipathic peptide), K-FGF, Ku70, Prion, pVEC, Pep-1, SynB1, Pep-7, HN-1, BGSC (Bis-Guanidinium-Spermidine-Cholesterol, and BGTC (Bis-Guanidinium-Tren-Cholesterol).

[151] In various embodiments, the cell-penetrating peptides of the present disclosure are cyclic cell-penetrating peptides (cCPPs). In some embodiment, CPPs are cyclized to form cCPP by forming a peptide bond between the N- and C-termini of two amino acids in a peptide sequence. In some embodiments, the cCPPs may include any combination of at least two arginines and at least two hydrophobic amino acids. In some embodiments, the cCPPs may include any combination of two to three arginines and at least two hydrophobic amino acids.

[152] In some embodiments, the cCPP used in compounds described herein has a structure comprising Formula III:



wherein:

each of AA₁, AA₂, AA₃, and AA₄, are independently selected from a D or L amino acid,

each of AA_u and AA_z, at each instance and when present, are independently selected from a D or L amino acid, and

m and n are independently selected from a number from 0 to 6; and

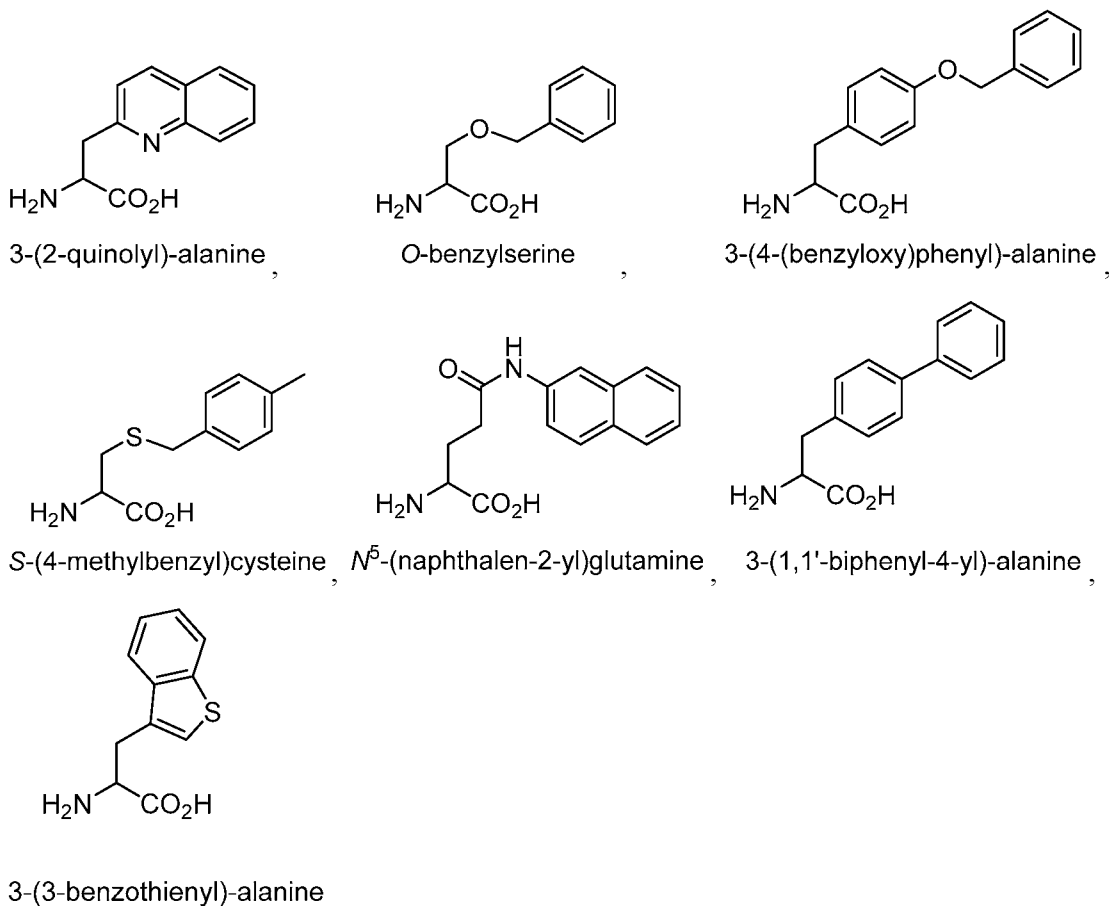
wherein:

at least two of AA_u, when present, AA₁, AA₂, AA₃, AA₄, and AA_z, when present, are independently arginine, and

at least two of AA_u, when present, AA₁, AA₂, AA₃, AA₄, and AA_z, when present, are independently a hydrophobic amino acid.

[153] In some embodiments, each hydrophobic amino acid is independently selected from is independently selected from glycine, alanine, valine, leucine, isoleucine, methionine, phenylalanine, tryptophan, proline, naphthylalanine, phenylglycine, homophenylalanine, tyrosine, cyclohexylalanine, piperidine-2-carboxylic acid, cyclohexylalanine, norleucine, 3-(3-benzothienyl)-alanine, 3-(2-quinolyl)-alanine, O-benzylserine, 3-(4-(benzyloxy)phenyl)-alanine, S-(4-methylbenzyl)cysteine, N-(naphthalen-2-yl)glutamine, 3-(1,1'-biphenyl-4-yl)-alanine, tert-leucine, or nicotinoyl lysine, each of which is optionally substituted with one or more substituents. The structures of certain of these non-natural aromatic hydrophobic amino acids (prior to incorporation into the peptides disclosed herein) are provided below. In particular embodiments, each hydrophobic amino acid is independently a hydrophobic

aromatic amino acid. In some embodiments, the aromatic hydrophobic amino acid is naphthylalanine, 3-(3-benzothienyl)-alanine, phenylglycine, homophenylalanine, phenylalanine, tryptophan, or tyrosine, each of which is optionally substituted with one or more substituents.



[154] The optional substituent can be any atom or group which does not significantly reduce (e.g., by more than 50%) the cytosolic delivery efficiency of the cCPP, e.g., compared to an otherwise identical sequence which does not have the substituent. In some embodiments, the optional substituent can be a hydrophobic substituent or a hydrophilic substituent. In certain embodiments, the optional substituent is a hydrophobic substituent. In some embodiments, the substituent increases the solvent-accessible surface area (as defined herein) of the hydrophobic amino acid. In some embodiments, the substituent can be a halogen, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, cycloalkynyl, heterocyclyl, aryl, heteroaryl, alkoxy, aryloxy, acyl, alkylcarbamoyl, alkylcarboxamidyl, alkoxycarbonyl, alkylthio, or arylthio. In some embodiments, the substituent is a halogen.

[155] Amino acids having higher hydrophobicity values can be selected to improve cytosolic delivery efficiency of a cCPP relative to amino acids having a lower hydrophobicity value. In

some embodiments, each hydrophobic amino acid independently has a hydrophobicity value which is greater than that of glycine. In other embodiments, each hydrophobic amino acid independently is a hydrophobic amino acid having a hydrophobicity value which is greater than that of alanine. In still other embodiments, each hydrophobic amino acid independently has a hydrophobicity value which is greater or equal to phenylalanine. Hydrophobicity may be measured using hydrophobicity scales known in the art. Table 2 below lists hydrophobicity values for various amino acids as reported by Eisenberg and Weiss (Proc. Natl. Acad. Sci. U. S. A. 1984;81(1):140–144), Engleman, et al. (Ann. Rev. of Biophys. Biophys. Chem.. 1986;1986(15):321–53), Kyte and Doolittle (J. Mol. Biol. 1982;157(1):105–132), Hoop and Woods (Proc. Natl. Acad. Sci. U. S. A. 1981;78(6):3824–3828), and Janin (Nature. 1979;277(5696):491–492), the entirety of each of which is herein incorporated by reference in its entirety. In particular embodiments, hydrophobicity is measured using the hydrophobicity scale reported in Engleman, et al.

Table 2.

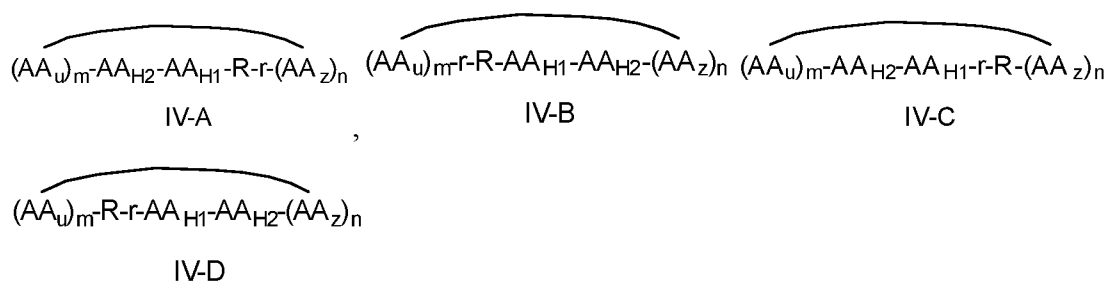
Amino Acid	Group	Eisenberg and Weiss	Engleman et al.	Kyrie and Doolittle	Hoop and Woods	Janin
Ile	Nonpolar	0.73	3.1	4.5	-1.8	0.7
Phe	Nonpolar	0.61	3.7	2.8	-2.5	0.5
Val	Nonpolar	0.54	2.6	4.2	-1.5	0.6
Leu	Nonpolar	0.53	2.8	3.8	-1.8	0.5
Trp	Nonpolar	0.37	1.9	-0.9	-3.4	0.3
Met	Nonpolar	0.26	3.4	1.9	-1.3	0.4
Ala	Nonpolar	0.25	1.6	1.8	-0.5	0.3
Gly	Nonpolar	0.16	1.0	-0.4	0.0	0.3
Cys	Unch/Polar	0.04	2.0	2.5	-1.0	0.9
Tyr	Unch/Polar	0.02	-0.7	-1.3	-2.3	-0.4
Pro	Nonpolar	-0.07	-0.2	-1.6	0.0	-0.3
Thr	Unch/Polar	-0.18	1.2	-0.7	-0.4	-0.2
Ser	Unch/Polar	-0.26	0.6	-0.8	0.3	-0.1
His	Charged	-0.40	-3.0	-3.2	-0.5	-0.1
Glu	Charged	-0.62	-8.2	-3.5	3.0	-0.7
Asn	Unch/Polar	-0.64	-4.8	-3.5	0.2	-0.5

Amino Acid	Group	Eisenberg and Weiss	Engleman et al.	Kyrie and Doolittle	Hoop and Woods	Janin
Gln	Unch/Polar	-0.69	-4.1	-3.5	0.2	-0.7
Asp	Charged	-0.72	-9.2	-3.5	3.0	-0.6
Lys	Charged	-1.10	-8.8	-3.9	3.0	-1.8
Arg	Charged	-1.80	-12.3	-4.5	3.0	-1.4

[156] The chirality of the amino acids can be selected to improve cytosolic uptake efficiency. In some embodiments, at least two of the amino acids have the opposite chirality. In some embodiments, the at least two amino acids having the opposite chirality can be adjacent to each other. In some embodiments, at least three amino acids have alternating stereochemistry relative to each other. In some embodiments, the at least three amino acids having the alternating chirality relative to each other can be adjacent to each other. In some embodiments, at least two of the amino acids have the same chirality. In some embodiments, the at least two amino acids having the same chirality can be adjacent to each other. In some embodiments, at least two amino acids have the same chirality and at least two amino acids have the opposite chirality. In some embodiments, the at least two amino acids having the opposite chirality can be adjacent to the at least two amino acids having the same chirality. Accordingly, in some embodiments, adjacent amino acids in the cCPP can have any of the following sequences: D-L; L-D; D-L-L-D; L-D-D-L; L-D-L-L-D; D-L-D-D-L; D-L-L-D-L; or L-D-D-L-D.

[157] In some embodiments, an arginine is adjacent to a hydrophobic amino acid. In some embodiments, the arginine has the same chirality as the hydrophobic amino acid. In some embodiments, at least two arginines are adjacent to each other. In still other embodiments, three arginines are adjacent to each other. In some embodiments, at least two hydrophobic amino acids are adjacent to each other. In other embodiments, at least three hydrophobic amino acids are adjacent to each other. In other embodiments, the cCPPs described herein comprise at least two consecutive hydrophobic amino acids and at least two consecutive arginines. In further embodiments, one hydrophobic amino acid is adjacent to one of the arginines. In still other embodiments, the cCPPs described herein comprise at least three consecutive hydrophobic amino acids and there consecutive arginines. In further embodiments, one hydrophobic amino acid is adjacent to one of the arginines. These various combinations of amino acids can have any arrangement of D and L amino acids, e.g., the sequences described above.

[158] In some embodiments, any four adjacent amino acids in the cCPPs described herein (e.g., the cCPPs according to Formula 2) can have one of the following sequences: AA_{H2}-AA_{H1}-R-r, AA_{H2}-AA_{H1}-r-R, R-r-AA_{H1}-AA_{H2}, or r-R-AA_{H1}-AA_{H2}, wherein each of AA_{H1} and AA_{H2} are independently a hydrophobic amino acid. Accordingly, in some embodiments, the cCPPs used in the compounds described herein comprise a structure according any of Formula IV-A-D:



wherein:

each of AA_{H1} and AA_{H2} are independently a hydrophobic amino acid;

at each instance and when present, each of AA_U and AA_Z are independently any amino acid; and

m and n are independently selected from a number from 0 to 6.

[159] In some embodiments, the total number of amino acids (including r, R, AA_{H1}, AA_{H2}), in the CPPs of Formula 4-A to 4-D are in the range of 6 to 10. In some embodiments, the total number of amino acids is 6. In some embodiments, the total number of amino acids is 7. In some embodiments, the total number of amino acids is 8. In some embodiments, the total number of amino acids is 9. In some embodiments, the total number of amino acids is 10.

[160] In some embodiments, the sum of m and n is from 2 to 6. In some embodiments, the sum of m and n is 2. In some embodiments, the sum of m and n is 3. In some embodiments, the sum of m and n is 4. In some embodiments, the sum of m and n is 5. In some embodiments, the sum of m and n is 6. In some embodiments, m is 0. In some embodiments, m is 1. In some embodiments, m is 2. In some embodiments, m is 3. In some embodiments, m is 4. In some embodiments, m is 5. In some embodiments, m is 6. In some embodiments, n is 0. In some embodiments, n is 1. In some embodiments, n is 2. In some embodiments, n is 3. In some embodiments, n is 4. In some embodiments, n is 5. In some embodiments, n is 6.

[161] In some embodiments, each hydrophobic amino acid is independently selected from glycine, alanine, valine, leucine, isoleucine, methionine, phenylalanine, tryptophan, proline, naphthylalanine, phenylglycine, homophenylalanine, tyrosine, cyclohexylalanine, piperidine-

2-carboxylic acid, or norleucine, each of which is optionally substituted with one or more substituents. In particular embodiments, each hydrophobic amino acid is independently a hydrophobic aromatic amino acid. In some embodiments, the aromatic hydrophobic amino acid is piperidine-2-carboxylic acid, naphthylalanine, phenylglycine, homophenylalanine, phenylalanine, tryptophan, or tyrosine, each of which is optionally substituted with one or more substituents. In particular embodiments, the hydrophobic amino acid is piperidine-2-carboxylic acid, naphthylalanine, tryptophan, or phenylalanine, each of which is optionally substituted with one or more substituents.

[162] In some embodiments, each of AA_{H1} and AA_{H2} are independently a hydrophobic amino acid having a hydrophobicity value which is greater than that of glycine. In other embodiments, each of AA_{H1} and AA_{H2} are independently a hydrophobic amino acid having a hydrophobicity value which is greater than that of alanine. In still other embodiments, each of AA_{H1} and AA_{H2} are independently an hydrophobic amino acid having a hydrophobicity value which is greater than that of phenylalanine, e.g., as measured using the hydrophobicity scales described above, including Eisenberg and Weiss (Proc. Natl. Acad. Sci. U. S. A. 1984;81(1):140–144), Engleman, et al. (Ann. Rev. of Biophys. Biophys. Chem. 1986; 1986(15):321–53), Kyte and Doolittle (J. Mol. Biol. 1982;157(1):105–132), Hoop and Woods (Proc. Natl. Acad. Sci. U. S. A. 1981;78(6):3824–3828), and Janin (Nature. 1979;277(5696):491–492), (see Table 1 above). In particular embodiments, hydrophobicity is measured using the hydrophobicity scale reported in Engleman, et al.

[163] The presence of a hydrophobic amino acid on the N- or C-terminal of a D-Arg or L-Arg, or a combination thereof, has also found to improve the cytosolic uptake of the cCPP (and the attached cargo). For example, in some embodiments, the cCPPs disclosed herein may include AA_{H1}-D-Arg or D-Arg-AA_{H1}. In other embodiments, the cCPPs disclosed herein may include AA_{H1}-L-Arg or L-Arg-AA_{H1}.

[164] The size of the hydrophobic amino acid on the N- or C-terminal of the D-Arg or an L-Arg, or a combination thereof (i.e., AA_{H1}), may be selected to improve cytosolic delivery efficiency of the CPP. For example, a larger hydrophobic amino acid on the N- or C-terminal of a D-Arg or L-Arg, or a combination thereof, improves cytosolic delivery efficiency compared to an otherwise identical sequence having a smaller hydrophobic amino acid. The size of the hydrophobic amino acid can be measured in terms of molecular weight of the hydrophobic amino acid, the steric effects of the hydrophobic amino acid, the solvent-accessible surface area (SASA) of the side chain, or combinations thereof. In some

embodiments, the size of the hydrophobic amino acid is measured in terms of the molecular weight of the hydrophobic amino acid, and the larger hydrophobic amino acid has a side chain with a molecular weight of at least about 90 g/mol, or at least about 130 g/mol, or at least about 141 g/mol. In other embodiments, the size of the amino acid is measured in terms of the SASA of the hydrophobic side chain, and the larger hydrophobic amino acid has a side chain with a SASA greater than alanine, or greater than glycine. In other embodiments, AA_{H1} has a hydrophobic side chain with a SASA greater than or equal to about piperidine-2-carboxylic acid, greater than or equal to about tryptophan, greater than or equal to about phenylalanine, or equal to or greater than about naphthylalanine. In some embodiments, AA_{H1} has a side chain side with a SASA of at least about 200 Å², at least about 210 Å², at least about 220 Å², at least about 240 Å², at least about 250 Å², at least about 260 Å², at least about 270 Å², at least about 280 Å², at least about 290 Å², at least about 300 Å², at least about 310 Å², at least about 320 Å², or at least about 330 Å². In some embodiments, AA_{H2} has a side chain side with a SASA of at least about 200 Å², at least about 210 Å², at least about 220 Å², at least about 240 Å², at least about 250 Å², at least about 260 Å², at least about 270 Å², at least about 280 Å², at least about 290 Å², at least about 300 Å², at least about 310 Å², at least about 320 Å², or at least about 330 Å². In some embodiments, the side chains of AA_{H1} and AA_{H2} have a combined SASA of at least about 350 Å², at least about 360 Å², at least about 370 Å², at least about 380 Å², at least about 390 Å², at least about 400 Å², at least about 410 Å², at least about 420 Å², at least about 430 Å², at least about 440 Å², at least about 450 Å², at least about 460 Å², at least about 470 Å², at least about 480 Å², at least about 490 Å², greater than about 500 Å², at least about 510 Å², at least about 520 Å², at least about 530 Å², at least about 540 Å², at least about 550 Å², at least about 560 Å², at least about 570 Å², at least about 580 Å², at least about 590 Å², at least about 600 Å², at least about 610 Å², at least about 620 Å², at least about 630 Å², at least about 640 Å², greater than about 650 Å², at least about 660 Å², at least about 670 Å², at least about 680 Å², at least about 690 Å², or at least about 700 Å². In some embodiments, AA_{H2} is a hydrophobic amino acid with a side chain having a SASA that is less than or equal to the SASA of the hydrophobic side chain of AA_{H1}. By way of example, and not by limitation, a cCPP having a Nal-Arg motif exhibits improved cytosolic delivery efficiency compared to an otherwise identical CPP having a Phe-Arg motif; a cCPP having a Phe-Nal-Arg motif exhibits improved cytosolic delivery efficiency compared to an otherwise identical cCPP having a Nal-Phe-Arg motif; and a phe-Nal-Arg motif exhibits improved cytosolic delivery efficiency compared to an otherwise identical cCPP having a nal-Phe-Arg motif.

[165] As used herein, “hydrophobic surface area” or “SASA” refers to the surface area (reported as square Ångstroms; Å²) of an amino acid side chain that is accessible to a solvent. In particular embodiments, SASA is calculated using the 'rolling ball' algorithm developed by Shrake & Rupley (*J Mol Biol.* 79 (2): 351–71), which is herein incorporated by reference in its entirety for all purposes. This algorithm uses a “sphere” of solvent of a particular radius to probe the surface of the molecule. A typical value of the sphere is 1.4 Å, which approximates to the radius of a water molecule.

[166] SASA values for certain side chains are shown below in Table 3. In certain embodiments, the SASA values described herein are based on the theoretical values listed in Table 3 below, as reported by Tien, et al. (PLOS ONE 8(11): e80635. <https://doi.org/10.1371/journal.pone.0080635>, which is herein incorporated by reference in its entirety for all purposes.

Table 3.

Residue	Theoretical	Empirical	Miller <i>et al.</i> (1987)	Rose <i>et al.</i> (1985)
Alanine	129.0	121.0	113.0	118.1
Arginine	274.0	265.0	241.0	256.0
Asparagine	195.0	187.0	158.0	165.5
Aspartate	193.0	187.0	151.0	158.7
Cysteine	167.0	148.0	140.0	146.1
Glutamate	223.0	214.0	183.0	186.2
Glutamine	225.0	214.0	189.0	193.2
Glycine	104.0	97.0	85.0	88.1
Histidine	224.0	216.0	194.0	202.5
Isoleucine	197.0	195.0	182.0	181.0
Leucine	201.0	191.0	180.0	193.1
Lysine	236.0	230.0	211.0	225.8
Methionine	224.0	203.0	204.0	203.4
Phenylalanine	240.0	228.0	218.0	222.8
Proline	159.0	154.0	143.0	146.8
Serine	155.0	143.0	122.0	129.8
Threonine	172.0	163.0	146.0	152.5
Tryptophan	285.0	264.0	259.0	266.3

Residue	Theoretical	Empirical	Miller <i>et al.</i> (1987)	Rose <i>et al.</i> (1985)
Tyrosine	263.0	255.0	229.0	236.8
Valine	174.0	165.0	160.0	164.5

[167] In some embodiments, the cCPP does not include a hydrophobic amino acid on the N- and/or C-terminal of AA_{H2}-AA_{H1}-R-r, AA_{H2}-AA_{H1}-r-R, R-r-AA_{H1}-AA_{H2}, or r-R-AA_{H1}-AA_{H2}. In alternative embodiments, the cCPP does not include a hydrophobic amino acid having a side chain which is larger (as described herein) than at least one of AA_{H1} or AA_{H2}. In further embodiments, the cCPP does not include a hydrophobic amino acid with a side chain having a surface area greater than AA_{H1}. For example, in embodiments in which at least one of AA_{H1} or AA_{H2} is phenylalanine, the cCPP does not further include a naphthylalanine (although the cCPP may include at least one hydrophobic amino acid which is smaller than AA_{H1} and AA_{H2}, e.g., leucine). In still other embodiments, the cCPP does not include a naphthylalanine in addition to the hydrophobic amino acids in AA_{H2}-AA_{H1}-R-r, AA_{H2}-AA_{H1}-r-R, R-r-AA_{H1}-AA_{H2}, or r-R-AA_{H1}-AA_{H2}.

[168] The chirality of the amino acids (i.e., D or L amino acids) can be selected to improve cytosolic delivery efficiency of the cCPP (and the attached cargo as described below). In some embodiments, the hydrophobic amino acid on the N- or C-terminal of an arginine (e.g., AA_{H1}) has the same or opposite chirality as the adjacent arginine. In some embodiments, AA_{H1} has the opposite chirality as the adjacent arginine. For example, when the arginine is D-arg (i.e., “r”), AA_{H1} is a D-AA_{H1}, and when the arginine is L-Arg (i.e., “R”), AA_{H1} is a L-AA_{H1}. Accordingly, in some embodiments, the cCPPs disclosed herein may include at least one of the following motifs: D-AA_{H1}-D-arg, D-arg-D-AA_{H1}, L-AA_{H1}-L-Arg, or L-Arg-LAA_{H1}. In particular embodiments, when arginine is D-arg, AA_{H1} can be D-nal, D-trp, or D-phe. In another non-limiting example, when arginine is L-Arg, AA_{H1} can be L-Nal, L-Trp, or L-Phe.

[169] In some embodiments, the cCPPs described herein include at least three arginines. Accordingly, in some embodiments, the cCPPs described herein include one of the following sequences: AA_{H2}-AA_{H1}-R-r-R, AA_{H2}-AA_{H1}-R-r-r, AA_{H2}-AA_{H1}-r-R-R, AA_{H2}-AA_{H1}-r-R-r, R-R-r-AA_{H1}-AA_{H2}, r-R-r-AA_{H1}-AA_{H2}, r-r-R-AA_{H1}-AA_{H2}, or, R-r-R-AA_{H1}-AA_{H2}. In particular embodiments, the cCPPs have one of the following sequences AA_{H2}-AA_{H1}-R-r-R, AA_{H2}-AA_{H1}-r-R-r, r-R-r-AA_{H1}-AA_{H2}, or R-r-R-AA_{H1}-AA_{H2}. In some embodiments, the chirality of AA_{H1} and AA_{H2} can be selected to improve cytosolic uptake efficiency, e.g., as described

above, where AA_{H1} has the same chirality as the adjacent arginine, and AA_{H1} and AA_{H2} have the opposite chirality.

[170] In some embodiments, the cCPPs described herein include three hydrophobic amino acids. Accordingly, in some embodiments, the cCPPs described herein include one of the following sequences: AA_{H3}-AA_{H2}-AA_{H1}-R-r, AA_{H3}-AA_{H2}-AA_{H1}-R-r, AA_{H3}-AA_{H2}-AA_{H1}-r-R, AA_{H3}-AA_{H2}-AA_{H1}-r-R, R-r-AA_{H1}-AA_{H2}-AA_{H3}, R-r-AA_{H1}-AA_{H2}-AA_{H3}, r-R-AA_{H1}-AA_{H2}-AA_{H3}, or, r-R-AA_{H1}-AA_{H2}-AA_{H3}, wherein AA_{H3} is any hydrophobic amino acid described above, e.g., piperidine-2-carboxylic acid, naphthylalanine, tryptophan, or phenylalanine. In some embodiments, the chirality of AA_{H1}, AA_{H2}, and AA_{H3} can be selected to improve cytosolic uptake efficiency, e.g., as described above, where AA_{H1} has the same chirality as the adjacent arginine, and AA_{H1} and AA_{H2} have the opposite chirality. In other embodiments, the size of AA_{H1}, AA_{H2}, and AA_{H3} can be selected to improve cytosolic uptake efficiency, e.g., as described above, where AA_{H3} has a SAS of less than or equal to AA_{H1} and/or AA_{H2}.

[171] In some embodiments, AA_{H1} and AA_{H2} have the same or opposite chirality. In certain embodiments, AA_{H1} and AA_{H2} have the opposite chirality. Accordingly, in some embodiments, the cCPPs disclosed herein include at least one of the following sequences: D-AA_{H2}-L-AA_{H1}-R-r; L-AA_{H2}-D-AA_{H1}-r-R; R-r-D-AA_{H1}-L-AA_{H2}; or r-R- L-AA_{H1}-D-AA_{H1}, wherein each of D-AA_{H1} and D-AA_{H2} is a hydrophobic amino acid having a D configuration, and each of L-AA_{H1} and L-AA_{H2} is a hydrophobic amino acid having an L configuration. In some embodiments, each of D-AA_{H1} and D-AA_{H2} is independently selected from the group consisting of D-pip, D-nal, D-trp, and D-phe. In particular embodiments, D-AA_{H1} or D-AA_{H2} is D-nal. In other particular embodiments, D-AA_{H1} is D-nal. In some embodiments, each of L-AA_{H1} and L-AA_{H2} is independently selected from the group consisting of L-Pip, L-Nal, L-Trp, and L-Phe. In particular embodiments, each of L-AA_{H1} and L-AA_{H2} is L-Nal. In other particular embodiments, L-AA_{H1} is L-Nal.

[172] As discussed above, the disclosure provides for various modifications to a cCPP which may improve cytosolic delivery efficiency. In some embodiments, improved cytosolic uptake efficiency can be measured by comparing the cytosolic delivery efficiency of the CPP having the modified sequence to a proper control sequence. In some embodiments, the control sequence does not include a particular modification (e.g., matching chirality of R and AA_{H1}) but is otherwise identical to the modified sequence. In other embodiments, the control has the following sequence: cyclic(FΦRRRRQ).

[173] As used herein cytosolic delivery efficiency refers to the ability of a cCPP to traverse a cell membrane and enter the cytosol. In embodiments, cytosolic delivery efficiency of the cCPP is not dependent on a receptor or a cell type. Cytosolic delivery efficiency can refer to absolute cytosolic delivery efficiency or relative cytosolic delivery efficiency.

[174] Absolute cytosolic delivery efficiency is the ratio of cytosolic concentration of a cCPP (or a cCPP-TP conjugate) over the concentration of the CPP (or the CPP-TP conjugate) in the growth medium. Relative cytosolic delivery efficiency refers to the concentration of a cCPP in the cytosol compared to the concentration of a control cCPP in the cytosol. Quantification can be achieved by fluorescently labeling the cCPP (e.g., with a FTIC dye) and measuring the fluorescence intensity using techniques well-known in the art.

[175] In particular embodiments, relative cytosolic delivery efficiency is determined by comparing (i) the amount of a CPP of the invention internalized by a cell type (e.g., HeLa cells) to (ii) the amount of the control CPP internalized by the same cell type. To measure relative cytosolic delivery efficiency, the cell type may be incubated in the presence of a cell-penetrating peptide of the invention for a specified period of time (e.g., 30 minutes, 1 hour, 2 hours, etc.) after which the amount of the CPP internalized by the cell is quantified using methods known in the art, e.g., fluorescence microscopy. Separately, the same concentration of the control cCPP is incubated in the presence of the cell type over the same period of time, and the amount of the control cCPP internalized by the cell is quantified.

[176] In other embodiments, relative cytosolic delivery efficiency can be determined by measuring the IC_{50} of a cCPP having a modified sequence for an intracellular target, and comparing the IC_{50} of the cCPP having the modified sequence to a proper control sequence (as described herein).

[177] In some embodiments, the relative cytosolic delivery efficiency of the cCPP-TP conjugates described herein is in the range of from about 1% to about 1000% compared to cyclo(F Φ RRRRQ), e.g., about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 100%, about 110%, about 120%, about 130%, about 140%, about 150%, about 160%, about 170%, about 180%, about 190%, about 200%, about 210%, about 220%, about 230%, about 240%, about 250%, about 260%, about 270%, about 280%, about 290%, about 300%, about 310%, about 320%, about 330%, about 340%, about 350%, about 360%, about 370%, about 380%, about 390%, about 400%, about 410%, about 420%, about 430%, about 440%, about 450%, about 460%, about 470%, about 480%, about 490%, about 500%, about 510%, about 520%, about 530%, about 540%, about

550%, about 560%, about 570%, about 580%, about 590%, about 600%, about 610%, about 620%, about 630%, about 640%, about 650%, about 660%, about 670%, about 680%, about 690%, about 700%, about 710%, about 720%, about 730%, about 740%, about 750%, about 760%, about 770%, about 780%, about 790%, about 800%, about 810%, about 820%, about 830%, about 840%, about 850%, about 860%, about 870%, about 880%, about 890%, about 900%, about 910%, about 920%, about 930%, about 940%, about 950%, about 960%, about 970%, about 980%, about 990%, about 1000%, inclusive of all values and subranges therebetween. cyclo(FΦRRRRQ).

[178] In other embodiments, the absolute cytosolic delivery efficacy of from about 40% to about 100%, e.g., about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, inclusive of all values and subranges therebetween.

[179] In some embodiments, the cCPP may be or include any of the sequences listed in Table 4. That is, the cCPPs used in the compounds disclosed herein may comprise any one of the sequences listed in Table 4, along with additional amino acids to form a cyclic sequence, or the sequences in the Table 4 may be cyclized (via a peptide bond) to form a cCPP. In some embodiments, the amino acids listed in Table 4 further include a glutamine residue or other amino acid that has a side chain that allows for conjugation of the TP.

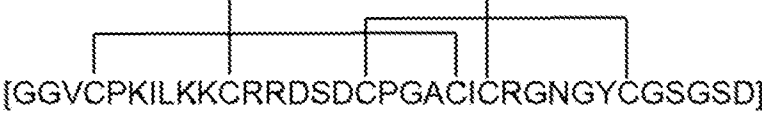
Table 4.

ID	Sequence
PCT 1	FΦRRR
PCT 2	FΦRRRC
PCT 3	FΦRRRU
PCT 4	RRRΦF
PCT 5	RRRRΦF
PCT 6	FΦRRRR
PCT 7	FφrRrR

PCT 8	FφrRrR
PCT 9	FΦRRRR
PCT 10	fΦRrRr
PCT 11	RRFRΦR
PCT 12	FRRRRΦ
PCT 13	rRFRΦR
PCT 14	RRΦFRR
PCT 15	CRRRRFW
PCT 16	FfΦRrRr
PCT 17	FFΦRRRR
PCT 18	RFRFRΦR
PCT 19	URRRRFW
PCT 20	CRRRRFW
PCT 21	FΦRRRRQK
PCT 22	FΦRRRRQC
PCT 23	fΦRrRrRQ
PCT 24	FΦRRRRRQ
PCT 25	RRRRΦFDΩC
PCT 26	FΦRRR
PCT 27	FWRRR

PCT 28	RRRΦF
PCT 29	RRRWF
SAR 1	FΦRRRR
SAR 19	FFRRR
SAR 20	FFrRr
SAR 21	FFRrR
SAR 22	FRFRR
SAR 23	FRRFR
SAR 24	FRRRF
SAR 25	GΦRRR
SAR 26	FFFRA
SAR 27	FFFRR
SAR 28	FFRRRR
SAR 29	FRRFRR
SAR 30	FRRRFR
SAR 31	RFFRRR
SAR 32	RFRRFR
SAR 33	FRFRRR
SAR 34	FFFRRR
SAR 35	FFRRRF
SAR 36	FRFFRR
SAR 37	RRFFFR
SAR 38	FFRFRR
SAR 39	FFRRFR
SAR 40	FRRFFR
SAR 41	FRRFRF
SAR 42	FRFRFR
SAR 43	RFFRFR
SAR 44	GΦRRRR
SAR 45	FFRRRRR
SAR 46	RFFRRRR

SAR 47	RRFFRRR
SAR 48	RFFFRRR
SAR 49	RRFFFRR
SAR 50	FFRRFRR
SAR 51	FFRRRRF
SAR 52	FRRFFRR
SAR 53	FFFRRRRR
SAR 54	FFFRRRRRR
SAR 55	FΦRrRr
SAR 56	XXRRRR
SAR 57	FfFRrR
SAR 58	fFfrRr
SAR 59	fFfRrR
SAR 60	FfFrRr
SAR 61	fFφrRr
SAR 62	fΦfrRr
SAR 63	φFfrRr
SAR 64	FΦrRr
SAR 65	fΦrRr
SAR 66	Ac-(Lys-fFRrRrD)
SAR 67	Ac-(Dap-fFRrRrD)
SAR 68	CWRRRRRC └S——S┐
SAR 69	CWWRRRRRC └S——S┐
SAR 70	CFWRRRRRC └S——S┐
SAR 71	CWWRRRRRC └S——S┐
Pin1 15	Pip-Nal-Arg-Glu-arg-arg-glu
Pin1 16	Pip-Nal-Arg-Arg-arg-arg-glu
Pin1 17	Pip-Nal-Nal-Arg-arg-arg-glu

Pin1 18	Pip-Nal-Nal-Arg-arg-arg-Glu
Pin1 19	Pip-Nal-Phe-Arg-arg-arg-glu
Pin1 20	Pip-Nal-Phe-Arg-arg-arg- Glu
Pin1 21	Pip-Nal-phe-Arg-arg-arg- glu
Pin1 22	Pip-Nal-phe-Arg-arg-arg- Glu
Pin1 23	Pip-Nal-nal-Arg-arg-arg- Glu
Pin1 24	Pip-Nal-nal-Arg-arg-arg- glu
Rev-13	[Pim-RQRR-Nlys]GRRR ^b
hLF	KCFQWQRNMRKVRGPPVSC
cTat	[KrRrGrKkRrE] ^c
cR10	[KrRrRrRrRrRE] ^c
L-50	[RVTRGKRRIRRpP]
L-51	[RTRTRGKRRIRVpP]
[WR]₄	[WRWRWRWR]
MCoTI-II	 [GGVCPKILKKCRRDSDCPGACICRGNGYCGSGSD]
Rotstein et al. <i>Chem. Eur. J.</i> 2011	[P-Cha-r-Cha-r-Cha-r-Cha-r-G] ^d
Lian et al. <i>J. Am. Chem. Soc.</i> 2014	Tm(SvP-F ₂ Pmp-H)-Dap-(FΦRRRR-Dap)] ^f
Lian et al. <i>J. Am. Chem. Soc.</i> 2014	[Tm(a-Sar-D-pThr-Pip-ΦRAa)-Dap-(FΦRRRR-Dap)] ^f
IA8b	[CRRSRGCGRRSRRCG] ^g
Dod-[R₅]	[K(Dod)RRRR]
LK-3	
	RRRR-[KRRRE] ^c
	RRR-[KRRRRE] ^c

	RR-[KRRRRRE] ^c
	R-[KRRRRRE] ^c
[CR] ₄	[CRCRCRCR]
cyc3	[Pra-LRKRLRKFRN-AzK] ^h
PMB	T-Dap-[Dap-Dap-f-L-Dap-Dap-T]
GPMB	T-Agp-[Dap-Agp-f-L-Agp-Agp-T]
cCPP1	FΦRRRR
cCPP12	FfΦRrRr
cCPP9	fΦRrRr
cCPP11	fΦRrRrR
cCPP18	FφrRrR
cCPP13	FφrRrR
cCPP6	FΦRRRRR
cCPP3	RRFRΦRQ
cCPP7	FFΦRRRR
cCPP8	RFRFRΦR
cCPP5	FΦRRR
cCPP4	FRRRRΦ
cCPP10	rRFRΦR
cCPP2	RRΦFRR
cCPP62	fΦfrRr

Φ, L-2-naphthylalanine; Pim, pimelic acid; Nlys, lysine peptoid residue; D-pThr, D-phosphothreonine; Pip, L-piperidine-2-carboxylic acid; Cha, L-3-cyclohexyl-alanine; Tm, trimesic acid; Dap, L-2,3-diaminopropionic acid; Sar, sarcosine; F₂Pmp, L-difluorophosphonomethyl phenylalanine; Dod, dodecanoyl; Pra, L-propargylglycine; AzK, L-6-Azido-2-amino-hexanoic; Agp, L-2-amino-3-guanidinypropionic acid; ^bCyclization between Pim and Nlys; ^cCyclization between Lys and Glu; ^dMacrocyclization by multicomponent reaction with aziridine aldehyde and isocyanide; ^eCyclization between the main-chain of Gln residue; ^fN-terminal amine and side chains of two Dap residues bicyclized with Tm; ^gThree Cys side chains bicyclized with tris(bromomethyl)benzene; ^hCyclization by the click reaction between Pra and Azk.

[180] Additionally, the cCPP used in the compounds and methods described herein can include any sequence disclosed in: U.S. App. No. 15/312,878; U.S. App. No. 15/360,719; International PCT Application Publication No. WO/2018/089648 (including the corresponding US publication), and International PCT Application Publication No. WO 2018/098231, each of which is incorporated by reference in its entirety for all purposes.

Thymidine Phosphorylase

[181] As discussed above, the compounds described herein include a wild type (wt) thymidine phosphorylase protein, or an active fragment or analog thereof (collectively referred to herein as “TP”). Thus, “TP” is used throughout the disclosure and the claims to refer to the wild type protein, or an active fragment or analog of wild type protein.

[182] As used herein, an “active fragment” refers to a portion of human or non-human wild type thymidine phosphorylase that exhibits an activity, such as one or more activities of a full-length thymidine phosphorylase or possesses another activity. In particular embodiments, a portion of wild type thymidine phosphorylase that shares at least one biological activity of wild type thymidine phosphorylase is considered to be an active fragment of thymidine phosphorylase. In some embodiments, the active fragment also includes at least one modification disclosed herein. Activity can be any percentage of activity (*i.e.*, more or less) of the full-length thymidine phosphorylase, including but not limited to, about 1% of the activity, about 2%, about 3%, about 4%, about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99%, about 100%, about 200%, about 300%, about 400%, about 500%, or more activity compared to the full-length thymidine phosphorylase. Thus, in some embodiments, the active fragment may be substituted for native thymidine phosphorylase and retain at least a portion of one or more biological activities of wild type thymidine phosphorylase. In other embodiments, the active fragment may be substituted for native thymidine phosphorylase and enhance one or more biological activities of wild type thymidine phosphorylase.

[183] The TP used in the present disclosure can be derived from any eukaryotic cell, e.g., mammalian cells. In some embodiments, the mammal is a mouse, human, bovine, rat, pig, horse, chicken, sheep, and the like. In particular embodiments, TP is human thymidine phosphorylase or derived from human thymidine phosphorylase. In some embodiments, TP is derived from *E. coli*.

[184] In some embodiments, TP for use in the compounds is full length human thymidine phosphorylase protein (SEQ ID NO. 1) or truncated protein (i.e., fragment), e.g., 1-10 propeptide-cleaved, (TP11; SEQ ID NO. 2), 1-15 peptide-cleaved (TP16; SEQ ID NO. 3) 1-21 peptide cleaved (TP21; SEQ ID NO 4), and 1-34 peptide cleaved (TP34; SEQ ID NO 5). In some embodiments, a suitable TP moiety may be a homologue or an analogue of truncated human thymidine phosphorylase or full-length human thymidine phosphorylase. For example, a homologue or an analogue of truncated or full-length human thymidine phosphorylase protein may be a modified thymidine phosphorylase protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring protein (e.g., SEQ ID NO. 1, NO. 2, NO. 3, NO. 4, or NO 5), while retaining substantial thymidine phosphorylase protein activity.

Table 5

Human Thymidine phosphorylase (TP) (P19971)

Full-Length (1-482)	MAALMTPGTGAPPAPGDFSGEGSQGLPDPSPEPKQLPELIRMKRDGGR LSEADIRGFVAAVVNGSAQGAQIGAMLMAIRLRGMDLEETSVLTQAL AQSGQQLEWPEAWRQQQLVDKHSTGGVGDKVSLVLAPALAACGCKV PMISGRGLGHTGGTLDKLESIPGFNVIQSPEQMQLLDQAGCCIVGQS EQLVPADGILYAARDVTATVDSLPLITASILSKKLVEGLSALVVDVKF GGAAVFPNQEQAARELAKTLVGVGASLGLRVAAALTAMDKPLGRCVG HALEVEEALLCMDGAGPPDLRDLVTTLGGALLWLSGHAGTQAQGAA RVAAALDDGSALGRFERMLAAQGVDPLARALCSGSPAERRQLLPR AREQEELLAPADGTVELVRALPLALVLHELGAAGRSRAGEPLRLGVGA ELLVDVGQRLRRGTPWLRVHRDGPALSGPQSRAEQEALVLSDRAPFA APSPFAELVLPPQQ (SEQ ID NO. 1)
Truncated (11-482) “TP11” (cleavage of 1-10 propeptide)	APPAPGDFSGEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVA AVVNGSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWP EAWRQQQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHT GGTLDKLESIPGFNVIQSPEQMQLLDQAGCCIVGQSEQLVPADGILY AARDVTATVDSLPLITASILSKKLVEGLSALVVDVKF GGAAVFPNQEQA ARELAKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLC MDGAGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGS ALGRFERMLAAQGVDPLARALCSGSPAERRQLLPRAREQEELLAPA DGTVELVRALPLALVLHELGAAGRSRAGEPLRLGVGAELLVDVGQR

	LRRGTPWLRVHRDGPALSGPQSRAALQEALVLSRAPFAAPSPFAELVLPPQQ (SEQ ID NO. 2)
Truncated (16-482) “TP16” (cleavage of 1-15 peptide)	GDFSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAVVN GSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEAWR QQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTL DKLESIPGFNVIQSPEQMQVLLDQAGCCIVGQSEQLVPADGILYAARD VTATVDSLPLITASILSKKLVEGLSALVVDVKFGGAAVFPNQEAREL AKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMDG AGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGR FERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGT VELVRALPLALVLHELGAGRSRAGEPLRLGVGAELLVDVGQRLRRGTP WLRVHRDGPALSGPQSRAALQEALVLSRAPFAAPSPFAELVLPPQQ (SEQ ID NO. 3)
Truncated (22-482) “TP22” (cleavage of 1-21 peptide)	GSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAVVNGSAQGA QIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEAWRQQLVDK HSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTL DKLESIPGFNVIQSPEQMQVLLDQAGCCIVGQSEQLVPADGILYAARD VTATVDSLPLITASILSKKLVEGLSALVVDVKFGGAAVFPNQEAREL AKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMDG AGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGR FERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGT VELVRALPLALVLHELGAGRSRAGEPLRLGVGAELLVDVGQR LRRGTPWLRVHRDGPALSGPQSRAALQEALVLSRAPFAAPSPFAELVL PPQQ (SEQ ID NO. 4)
Truncated (35-482) “TP35” (cleavage of 1-34 peptide)	QLPELIRMKRDGGRLSEADIRGFVAAVVNGSAQGAQIGAMLMAIRLR GMDLEETSVLTQALAQSGQQLEWPEAWRQQLVDKHSTGGVGDKVS LVLAPALAACGCKVPMISGRGLGHTGGTL DKLESIPGFNVIQSPEQMQVLLDQAGCCIVGQSEQLVPADGILYAARD VTATVDSLPLITASILSKKLVEGLSALVVDVKFGGAAVFPNQEAREL AKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMDG AGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGR FERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGT VELVRALPLALVLHELGAGRSRAGEPLRLGVGAELLVDVGQR

	LRRGTPWLRVHRDGPALSGPQSRAEQEALVLSRAPFAAPSPFAELVL PPQQ (SEQ ID NO. 5)
--	--

[185] Thus, in some embodiments, TP suitable for the compounds and methods described herein is substantially homologous to full-length human thymidine phosphorylase protein (SEQ ID NO. 1). In some embodiments, the TP suitable for the present compounds and methods has an amino acid sequence that is at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91 %, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or more homologous to SEQ ID NO. 1. In some embodiments, the TP suitable for the compounds and methods described herein has an amino acid sequence 95% or more homologous to SEQ ID NO:1.

[186] In some embodiments, TP suitable for the compounds and methods described herein is substantially homologous to truncated (e.g., 1-10, 1-15, 1-21, or 1-34 propeptide-cleaved) thymidine phosphorylase protein (SEQ ID NO: 2, 3, 4, or 5). In some embodiments, TP suitable for the present compounds and methods has an amino acid sequence at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91 %, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or more homologous to SEQ ID NO. 2, 3, 4, 5 or 5. In some embodiments, the TP suitable for the compounds and methods described herein has an amino acid sequence 95% or more homologous to SEQ ID NO:2, 3, 4, or 5.

[187] Thus, in some embodiments, TP suitable for the compounds and methods described herein is substantially identical to full-length human TP protein (SEQ ID NO. 1). In some embodiments, TP suitable for the present compounds and methods herein has an amino acid sequence at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91 %, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or more identical to SEQ ID NO:1. In some embodiments, the TP suitable for the compounds and methods described herein has an amino acid sequence 95% or more identical to SEQ ID NO:1.

[188] Thus, in some embodiments, TP suitable for the compounds and methods described herein is substantially identical to truncated (1-10 propeptide-cleaved) human TP protein (SEQ ID NO. 2). In some embodiments, TP suitable for the present compounds and methods has an amino acid sequence at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91 %, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or more identical to SEQ ID NO: 2, 3, 4, or 5. In some embodiments, the TP suitable for the compounds and methods described herein has an amino acid sequence 95% or more identical to SEQ ID NO: 2, 3, 4, or 5.

[189] In some embodiments, the TP suitable for the compounds and methods described herein contains a fragment of full-length human TP protein (SEQ ID NO. 1) or a fragment of truncated (1-10, 1-15, 1-21, or 1-34 propeptide-cleaved) thymidine phosphorylase protein (SEQ ID NO: 2, 3, 4, or 5).

[190] As discussed above, TP can be the wild-type human protein or an active fragment of wild type human protein which can be substituted for native thymidine phosphorylase. In some embodiments, an active fragment of thymidine phosphorylase can rescue one or more phenotypes or symptoms associated with MNGIE or symptoms associated with thymidine phosphorylase-deficiency once located to the cytosol. An active fragment of the wild-type sequence is a sequence which functions in a substantially similar manner to the wild-type protein. Thus, the active fragment includes any amino acid sequence of the wild-type protein that, when located in the cytosol, allows the cell to function substantially similar to a similar cell which otherwise includes wild-type TP. In some embodiments, the active fragment includes an amino acid sequence which results in an insignificant decrease in function after cytosol entry compared to the wild-type TP but still exhibits the desired therapeutic effect, e.g., about 1%, about 2%, about 3%, about 4%, about 5%, about 6%, about 7%, about 8%, about 9%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, or about 50%.

[191] In some embodiments, the active fragment of the wild-type protein may have amino acid sequence that is reduced by about 1 or more amino acids, e.g., about 5, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, about 70, about 75, about 80, about 90, about 95, about 100, about 105, about 110,

about 115, about 120, about 125, about 130, about 135, about 140, about 145, about 150, about 155, about 160, about 165, about 170, about 175, or about 180 or more amino acids.

[192] As used herein, an “analog” refers to a variant of TP which has phosphorylase activity, but one or more properties of the variant are improved relative to wild-type TP. For example, the phosphorylase activity can be improved to enhancing binding and/or enzymatic activity through protein engineering, or stability may be enhanced, either through protein engineering or conjugation of a water-soluble polymer, e.g., as described herein. In other embodiments, one or more properties (other than the phosphorylase activity) of the wild-type TP are either not present (eliminated) or are reduced in the “analog.” Non-limiting examples of properties that may be reduced or eliminated include immunogenic, angiogenic, thrombogenic, and SRC homology 3 domain (SH3 domain) binding activity. For example, the interaction between a PXXP sequence and the SH3 domain on certain proteins, such as between the PXXP sequence on Fyn and the SH3 domain on Lyn, is believed to increase the risk of thrombosis (Circ Res. 2014, 115(12): 997-1006). Amino acids 12-15 (PPAP) in wild-type human TP have this sequence. Thus truncated TP sequences (SEQ. NO. 3, NO. 4, or NO. 5) with this PPAP sequence removed were designed.

[193] In some such embodiments, TP may be fused or conjugated to a moiety that improves half-life or stability. Non-limiting examples of such moieties include proteins and water-soluble polymers. In some embodiments, TP may be fused or conjugated to Fc or human serum albumin (HSA). Without being bound by theory, Fc and HSA interact with FCRn receptor, and this activity allows for Fc and/or HSA fusions with TP to increase the circulating half-life of TP. In some embodiments, Fc or HSA is located on the N-terminal of TP, with or without the PPAP sequence (amino acids 12-15). In other embodiments, Fc or HSA is located on the C-terminal of TP, with or without the PPAP sequence (amino acids 12-15). In some embodiments, a linker can be used to connect TP (with or with PPAP) to Fc or HSA. Non-limiting examples of such constructs are provided below in Table 6. The present disclosure contemplates sequences having a sequence identity of at least about 85 % of those provided in Table 6

Table 6
Example TP and Fc Conjugates

IgG1Fc-TP (16-482) N-terminal fusion	EPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVTV DVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREE MTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSF FLYSKLTVDKSRWQQGNVSCSVMHEALHNHYTQKSLSLSPGKGDFS GEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAVVNGSAQ GAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEAWRQQLV DKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTLDKLES IPGFNVISPEQMQLVLLDQAGCCIVGQSEQLVPADGILYAARDVTATV DSLPLITASILSKKLVEGLSALVVDVKFGGA AVFPNQEARELAKTLV GVGASLGLRVAAALTAMDKPLGRCVGHAEVVEEALLCMDGAGPPDL RDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGRFERML AAQGVDPGLARALCSGSPAERRQLPRAREQEELLAPADGTVELVRA LPLALVLHELGAGRSRAGEPLRLGVGAELLVDVGQRLRRGTPWLRVH RDGPALSGPQSRALQEALVLSRAPFAAPSPFAELVLPPQQ (SEQ ID NO. 6)
--	---

IgG1Fc-TP (16-482) N-terminal fusion with GGGS linker “Fc-TP16”	EPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVTV DVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREE MTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSF FLYSKLTVDKSRWQQGNVFSVSMHEALHNHYTQKSLSLSPGKGGG GSGDFSGEGSQGLPDPSPPEPKQLPELIRMKRDGGRLSEADIRGFVAAV VNGSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEA WRQQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGG TLDKLESIPGFNVIQSPEQMQLLDQAGCCIVGQSEQLVPADGILYAA RDVTATVDSLPLITASILSKKLVEGLSALVVDVKFGGA AVFPNQEQAR ELAKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMD GAGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALG RFERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGT VELVRALPLALVLHELGAGRSRAGEPLRLGVGAELLVDVGQRLRRGT PWLRVHRDGPALSGPQSRALQEALVLSRAPFAAPSPFAELVLPPQQ (SEQ ID NO. 7)
--	---

TP (16-482)- IgG1Fc C-terminal fusion	GDFSGEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAVVN GSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEAWR QQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTL DKLESIPGFNVIQSPEQMQLVLLDQAGCCIVGQSEQLVPADGILYAARD VTATVDSLPLITASILSKKLVEGLSALVVDVKFGGAAVFPNQEQAREL AKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMDG AGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGR FERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGTV ELVRALPLALVLHELGAAGRSRAGEPLRLGVGAELLVDVGQRLRRGTP WLRVHRDGPALSGPQSRALQEALVLSDRAPFAAPSPFAELVLPPQQEP KSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDV SHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQD WLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTK NQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYS KLTVDKSRWQQGNVFSQSVMHAEALHNHYTQKSLSLSPGK (SEQ ID NO. 8)
--	---

TP (16-482)- IgG1Fc C-terminal fusion with GGGS linker “TP16-Fc”	GDFSGEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAVVN GSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEAWR QQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTL DKLESIPGFNVIQSPEQMQLLDQAGCCIVGQSEQLVPADGILYAARD VTATVDSLPLITASILSKKLVEGLSALVVDVKFGGAAVFPNQEQAREL AKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMDG AGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGR FERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGTV ELVRALPLALVLHEL GAGRSRAGEPLRLGVGAELLVDVGQRLRRGTP WLRVHRDGPALSGPQSRALQEALVLSDRAPFAAPSPFAELVLPPQQG GGGSEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVT CVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVL TVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPS REEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSD GSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (SEQ ID NO. 9)
--	---

TP (16-482)- IgG1Fc_R435 C-terminal fusion with GGGS linker	GDFS GEGSQGLPDPSPEPKQLPELIRMKR DGGRLSEADIRGFVAAVVN GSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEAWR QQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTL DKLESIPGFNVIQSPEQM QVLLDQAGCCIVGQSEQLVPADGILYAARD VTATVDSLPLITASILSKKLVEGLSALVVDVKFGGA AVFPNQEQAREL AKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMDG AGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGR FERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGTV ELVRALPLALVLHEL GAGRSRAGEPLRLGVGAELLVDVGQRLRRGTP WLRVHRDGPALSGPQSRALQEALVLSRAPFAAPSPFAELVLPPQQG GGGSEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVT CVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVL TVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPS REEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSD GSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNRYTQKSLSLSPGK (SEQ ID NO. 10)
--	---

IgG1Fc(R435)-TP (16-482) N-terminal fusion with GGGGS linker	EPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVTVV DVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREE MTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSF FLYSKLTVDKSRWQQGNVFSVSMHEALHNRYTQKSLSLSPGKGGG GSGDFSGEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAV VNGSAQGAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEA WRQQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGG TLDKLESIPGFNVIQSPEQMQLLDQAGCCIVGQSEQLVPADGILYAA RDVTATVDSLPLITASILSKKLVEGLSALVVDVKFGGA AVFPNQE QAR ELAKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMD GAGPPDLRDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALG RFERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGT VELVRALPLALVLHELGAGRSRAGEPLRLGVGAELLVDVGQRLRRGT PWLRVHRDGPALSGPQSRALQEALVLSRAPFAAPSPFAELVLPPQQ (SEQ ID NO. 11)
--	---

IgG1Fc- TP_R435 (16-482) N-terminal fusion	EPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVTV DVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREE MTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSF FLYSKLTVDKSRWQQGNVSCSVMHEALHNRYTQKSLSPGKGDFA GEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAVVNGSAQ GAQIGAMLMAIRLRGMDLEETSVLTQALAQSGQQLEWPEAWRQQLV DKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTLDKLES IPGFNVISPEQMQLLDQAGCCIVGQSEQLVPADGILYAARDVTATV DSLPLITASILSKKLVEGLSALVVDVKFGGA AVFPNQEARELAKTLV GVGASLGLRVAAALTAMDKPLGRCVGHAEVVEEALLCMDGAGPPDL RDLVTTLGGALLWLSGHAGTQAQGAARVAAALDDGSALGRFERML AAQGVDPGLARALCSGSPAERRQLPRAREQEELLAPADGTVELVRA LPLALVLHELGAGRSRAGEPLRLGVGAELLVDVGQRLRRGTPWLRVH RDGPALSGPQSRALQEALVLSRAPFAAPSPFAELVLPPQQ (SEQ ID NO. 12)
--	--

TP (16-482) - IgG1Fc_R435 C-terminal fusion	GDFSSEGSQGLPDPSPEPKQLPELIRMKRDGGRLSEADIRGFVAAVVN GSAQGAQIGAMLMAILRGMDLEETSVLTQALAQSGQQLEWPEAWR QQLVDKHSTGGVGDKVSLVLAPALAACGCKVPMISGRGLGHTGGTL DKLESIPGFNVIQSPEQMQLLDQAGCCIVGQSEQLVPADGILYAARD VTATVDSLPLITASILSKKLVEGLSALVVDVKFGGAAVFPNQEQAREL AKTLVGVGASLGLRVAAALTAMDKPLGRCVGHAEVEEALLCMDG AGPPDLRDLVTTLGALLWLSGHAGTQAQGAARVAAALDDGSALGR FERMLAAQGVDPGLARALCSGSPAERRQLLPRAREQEELLAPADGTV ELVRALPLALVLHELGAAGRSRAGEPLRLGVGAELLVDVGQRLRRGTP WLRVHRDGPALSGPQSRALQEALVLSRAPFAAPSPFAELVLPPQQEP KSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDV SHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQD WLNQKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTK NQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYS KLTVDKSRWQQGNVFSVSMHEALHNRYTQKSLSLSPGK (SEQ ID NO. 13)
--	---

[194] In some embodiments, TP in the present compounds and methods has an amino acid sequence at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91 %, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or more identical to SEQ ID NO: 6, 7, 8, 9, 10, 11, 12, or 13. In some embodiments, the TP suitable for the compounds and methods described herein has an amino acid sequence 95% or more identical to SEQ ID NO: 6, 7, 8, 9, 10, 11, 12, or 13.

[195] In some embodiments, the TP analog contains one or more amino acid substitutions. Skilled persons can use molecular modeling to select mutations that are likely to be structurally tolerated, *e.g.* deletion in loops, insertion in loops, deletion of domains, C-terminal truncations, and N-terminal truncations. Homology modeling against TP variants from other organisms may be used to identify amino-acid residues as tolerant of mutations. Modeling is also used to select mutations that alter the function of the enzyme, such as mutations in and near the active

site of the enzyme. In some embodiments, the substitutions may be conservative substitutions or non-conservative substitutions.

[196] Examples of conservative amino acid substitutions include substitution of one amino acid for another amino acid within one from one of the following groups: basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). In some embodiments, structurally similar amino acids are substituted to reverse the charge of a residue (*e.g.*, glutamine for glutamic acid or vice-versa, aspartic acid for asparagine or vice-versa). In some embodiments, tyrosine is substituted for phenylalanine or vice-versa. Other non-limiting examples of amino acid substitutions are described, for example, by H. Neurath and R. L. Hill, 1979, *In, The Proteins*, Academic Press, New York. Common substitutions are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly.

[197] In some embodiments, the TP may be conjugated to a pharmaceutically acceptable water soluble polymer. Non-limiting examples of pharmaceutically acceptable water soluble polymers include polyethylene glycol (PEG), dextran or poly(n-vinyl pyrrolidone)polyethylene glycol, propylene glycol homopolymers, polypropylene oxide/ethylene oxide co-polymers, polyoxyethylated polyols, polyvinyl alcohol, polyoxyethylated polyols, polyoxyethylated sorbitol, polyoxyethylated glucose, polyoxyethylated glycerol (POG), polyoxyalkylenes, polyethylene glycol propionaldehyde, copolymers of ethylene glycol/propylene glycol, monomethoxy-polyethylene glycol, mono-(C1-C10) alkoxy- or aryloxy-polyethylene glycol, carboxymethylcellulose, polyacetals, polyvinyl alcohol (PVA), polyvinyl pyrrolidone, poly-1,3-dioxolane, poly-1,3,6-trioxane, ethylene/maleic anhydride copolymer, poly(β -amino acids) (either homopolymers or random copolymers), poly(n-vinyl pyrrolidone)polyethylene glycol, propylene glycol homopolymers (PPG) and other polyalkylene oxides, polypropylene oxide/ethylene oxide copolymers, colonic acids or other polysaccharide polymers, Ficoll or and mixtures thereof. In particular embodiments, the TP is a PEGylated. As used herein “PEGylation” refers to the coupling of TP to one or more polyethylene glycol (PEG) residues. In some embodiments, the molecular weight of the PEG is from about 0.1 kDa to about 100 kDa, *e.g.*, about 0.1 kDa, about 1 kDa, about 10 kDa, about 20 kDa, about 30 kDa, about 40 kDa, about 50 kDa, about

60 kDa, about 70 kDa, about 80 kDa, about 90 kDa, and about 100 kDa. In particular embodiments, the PEG is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 kDa, including 2 kDa or about 5 kDa. The polymer can be linear or branched. The attachment of the such polymers (*e.g.* PEG) adds molecular weight to the TP and may lead to an increased half-life by improving stability, and/or reducing degradation and/or excretion. Conjugation of the polymers may also improve the solubility and stability in aqueous solutions at physiological pH while retaining biological activity of TP. PEG, and any other biological polymers, can be attached to HPPD at any suitable site, *e.g.*, the N- or C-termini, or the side chain of any amino acid which has a functional group suitable for conjugate or which can be synthetically modified.

[198] The above polymers, such as PEG groups, can be attached to the TP under any suitable conditions used to react a protein with an activated polymer molecule. Any means known in the art can be used, including via acylation, reductive alkylation, Michael addition, thiol alkylation or other chemoselective conjugation/ligation methods through a reactive group on the PEG moiety (*e.g.*, an aldehyde, amino, ester, thiol, α -haloacetyl, maleimido or hydrazino group) to a reactive group on the TP (*e.g.*, an aldehyde, amino, ester, thiol, α -haloacetyl, maleimido or hydrazino group). Activating groups which can be used to link the water soluble polymer to one or more proteins include without limitation sulfone, maleimide, sulfhydryl, thiol, triflate, tresylate, aziridine, oxirane, 5-pyridyl, and alpha-halogenated acyl group (*e.g.*, α -iodo acetic acid, α -bromoacetic acid, α -chloroacetic acid). If attached to the TP by reductive alkylation, the polymer selected should have a single reactive aldehyde so that the degree of polymerization is controlled. See, for example, Kinstler et al., *Adv. Drug. Delivery Rev.* 54: 477-485 (2002); Roberts et al., *Adv. Drug Delivery Rev.* 54: 459-476 (2002); and Zalipsky et al., *Adv. Drug Delivery Rev.* 16: 157-182 (1995).

[199] The TP can be linked to the above polymers via direct covalent linkage by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C-terminal residues of these targeted amino acids. Reactive groups on the peptide or conjugate moiety include, *e.g.*, an aldehyde, amino, ester, thiol, α -haloacetyl, maleimido or hydrazino group. Derivatizing agents include, for example, maleimidobenzoyl sulfosuccinimide ester (conjugation through cysteine residues), N-hydroxysuccinimide (through lysine residues), glutaraldehyde, succinic anhydride or other agents known in the art. Alternatively, the conjugate moieties can be linked to the TP indirectly through intermediate carriers, such as polysaccharide or polypeptide carriers. Examples of

polysaccharide carriers include aminodextran. Examples of suitable polypeptide carriers include polylysine, polyglutamic acid, polyaspartic acid, co-polymers thereof, and mixed polymers of these amino acids and others, e.g., serines, to confer desirable solubility properties on the resultant loaded carrier.

[200] In embodiments, a thiol moiety within a TP is modified with a water-soluble polymer, such as PEG. In some embodiments, the thiol is modified with maleimide-activated PEG in a Michael addition reaction to result in a PEGylated peptide comprising the thioether linkage. In alternative embodiments, a thiol is modified with a haloacetyl-activated PEG in a nucleophilic substitution reaction to result in a PEGylated peptide comprising the thioether linkage. Cysteinyll residues are most commonly reacted with α -haloacetates (and corresponding amines), such as chloroacetic acid and chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyll residues also are derivatized by reaction with bromotrifluoroacetone, α -bromo- β -(5-imidazolyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

[201] Histidyl residues are derivatized by reaction with diethylpyrocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

[202] Lysinyll and amino-terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyll residues. Other suitable reagents for derivatizing α -amino-containing residues include imidoesters such as methyl picolinimidate, pyridoxal phosphate, pyridoxal, chloroborohydride, trinitrobenzenesulfonic acid, O-methylisourea, 2,4-pentanedione, and transaminase-catalyzed reaction with glyoxylate.

[203] Arginyll residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pKa of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group.

[204] The specific modification of tyrosyl residues may be made, with particular interest in introducing spectral labels into tyrosyl residues by reaction with aromatic diazonium

compounds or tetranitromethane. Most commonly, N-acetylimidazole and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

[205] Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimides (R--N.dbd.C.dbd.N--R'), where R and R' are different alkyl groups, such as 1-cyclohexyl-3-(2-morpholinyl-4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

[206] Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (T. E. Creighton, *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), deamidation of asparagine or glutamine, acetylation of the N-terminal amine, and/or amidation or esterification of the C-terminal carboxylic acid group.

[207] In another embodiment, the present invention relates to variants of the polypeptide of SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, or any other sequence disclosed herein, comprising a substitution, deletion, and/or insertion at one or more (*e.g.*, several) positions. In an embodiment, the number of amino acid substitutions, deletions and/or insertions introduced into the mature polypeptide of SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, is not more than 50, or not more than 40, or not more than 30, or not more than 20, or not more than 10, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. The amino acid changes may be of a minor nature, that is conservative amino acid substitutions or insertions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of 1-30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

[208] Alternatively, the amino acid changes may be of such a nature that the physico-chemical properties of the polypeptides are altered. For example, amino acid changes may improve the thermal stability of the polypeptide, alter the substrate specificity, change the pH optimum, and the like.

[209] Essential amino acids in a polypeptide can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, 1989, *Science* 244: 1081-1085). In the latter technique, single alanine mutations are

introduced at every residue in the molecule, and the resultant mutant molecules are tested for activity comparable to native thymidine phosphorylase to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton et al., 1996, *J. Biol. Chem.* 271: 4699-4708. The active site of the enzyme or other biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction, or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos et al., 1992, *Science* 255: 306-312; Smith et al., 1992, *J. Mol. Biol.* 224: 899-904; Wlodaver et al., 1992, *FEBS Lett.* 309: 59-64. The identity of essential amino acids can also be inferred from an alignment with a related polypeptide.

[210] Single or multiple amino acid substitutions, deletions, and/or insertions can be made and tested using known methods of mutagenesis, recombination, and/or shuffling, followed by a relevant screening procedure, such as those disclosed by Reidhaar-Olson and Sauer, 1988, *Science* 241: 53-57; Bowie and Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86: 2152-2156; WO 95/17413; or WO 95/22625. Other methods that can be used include error-prone PCR, phage display (e.g., Lowman et al., 1991, *Biochemistry* 30: 10832-10837; U.S. Pat. No. 5,223,409; WO 92/06204), and region-directed mutagenesis (Derbyshire et al., 1986, *Gene* 46: 145; Ner et al., 1988, *DNA* 7: 127).

[211] Non-limiting examples of the peptide conjugates disclosed herein are provided in the following table.

cCPP	Linker (L)	TP	Water-soluble polymer
cCPP9	PEG-12	TP	PEG10K
	PEG-8	TP	PEG5K
	PEG-24	TP	PEG40K
	PEG-4	TP	PEG10K
	PEG-12	TP11	PEG10K
	PEG-8	TP11	PEG5K
	PEG-24	TP11	PEG40K
	PEG-4	TP11	PEG10K
	PEG-12	TP16	PEG10K
	PEG-8	TP16	PEG5K
	PEG-24	TP16	PEG40K

	PEG-4	TP16	PEG10K
	PEG-12	TP35	PEG10K
	PEG-8	TP35	PEG5K
	PEG-24	TP35	PEG40K
cCPP11	PEG-12	TP	PEG10K
	PEG-8	TP	PEG5K
	PEG-24	TP	PEG40K
	PEG-4	TP	PEG10K
	PEG-12	TP11	PEG10K
	PEG-8	TP11	PEG5K
	PEG-24	TP11	PEG40K
	PEG-4	TP11	PEG10K
	PEG-12	TP16	PEG10K
	PEG-8	TP16	PEG5K
	PEG-24	TP16	PEG40K
	PEG-4	TP16	PEG10K
	PEG-12	TP35	PEG10K
	PEG-8	TP35	PEG5K
	PEG-24	TP35	PEG40K
cCPP12	PEG-12	TP	PEG10K
	PEG-8	TP	PEG5K
	PEG-24	TP	PEG40K
	PEG-4	TP	PEG10K
	PEG-12	TP11	PEG10K
	PEG-8	TP11	PEG5K
	PEG-24	TP11	PEG40K
	PEG-4	TP11	PEG10K
	PEG-12	TP16	PEG10K
	PEG-8	TP16	PEG5K
	PEG-24	TP16	PEG40K
	PEG-4	TP16	PEG10K
	PEG-12	TP35	PEG10K

	PEG-8	TP35	PEG5K
	PEG-24	TP35	PEG40K
	PEG-4	TP35	PEG10K

[212] In some embodiments, cCPP may be conjugated, via the linker, to the N or C terminus of the TP. In some embodiments, the linker further comprises an amino acid (e.g., lysine), which to facilitate chemical conjugation of the TP to a side chain of an amino acid on the cCCP

[213] In some embodiments, the water-soluble polymer can be conjugated to any suitable amino acid side chain in TP, e.g., lysine, glutamine, glutamic acid, asparagine, aspartic acid, and the like.

Methods of Treatment

[214] In embodiments of the present disclosure, a method of treating Mitochondrial Neurogastrointestinal Encephalopathy in a patient in need thereof, comprising administering a compound disclosed herein is provided.

[215] MNGIE impacts both the digestive systems and nervous system of patients afflicted with this disease. In various embodiments, treatment therefore refers to partial or complete alleviation, amelioration, relief, inhibition, delaying onset, reducing severity and/or incidence of digestive and nervous system impairment of a patient. In other embodiments, treatment therefore refers to partial or complete alleviation, amelioration, relief, inhibition, delaying onset, reducing severity and/or incidence of digestive system impairment of a patient. As used herein, the term “digestive system impairment” includes various symptoms associated with impairment of the gastrointestinal system. Symptoms of digestive system impairment may include, for example, gastrointestinal dysmotility in which the muscles and nerves of the digestive system do not move food through the digestive tract efficiently. The resulting digestive problems include feelings of fullness (satiety) after eating only a small amount, trouble swallowing (dysphagia), nausea and vomiting after eating, episodes of abdominal pain, diarrhea, and intestinal blockage. These gastrointestinal conditions lead to extreme weight loss and reduced muscle mass (cachexia). In some embodiments, treatment refers to partial or complete alleviation, relief, inhibition, delaying onset, reducing severity and/or incidence of gastrointestinal dysmotility and the accompanying conditions.

[216] MNGIE disease is also characterized by abnormalities of the nervous system. Affected individuals can experience tingling, numbness, and weakness in their limbs (peripheral neuropathy), particularly in the hands and feet. Additional neurological signs and symptoms

can include droopy eyelids (ptosis), weakness of the muscles that control eye movement (ophthalmoplegia), and hearing loss. Leukoencephalopathy, which is the deterioration of a type of brain tissue known as white matter, is a hallmark of MNGIE disease. In various embodiments, treatment therefore refers to partial or complete alleviation, amelioration, relief, inhibition, delaying onset, reducing severity and/or incidence of nervous system impairment of a patient, including but not limited to conditions such as ptosis, ophthalmoplegia, and hearing loss. In related embodiments, the methods of treatment provide partial or complete alleviation, amelioration, relief, inhibition, delaying onset, reducing severity and/or incidence of the peripheral neuropathy that can be a neurological symptom of MNGIE.

[217] In some embodiments, a method is provided for reducing extracellular and intracellular levels of thymidine in a patient in need thereof, comprising administering a compound disclosed herein. That is, not only does intracellular delivery of TP as described here reduce intracellular levels of thymidine, but it also reduces extracellular levels of thymidine in circulation. Mutations in *TYMP*, the gene that provides instructions for making TP, are believed to cause MNGIE disease by reducing or eliminating appropriate levels of enzymatic activity of this protein. Excess levels of thymidine that can result from these mutations are toxic to the body, leading to the disruption of the usual maintenance and repair of mitochondrial DNA. Without being bound by theory, the resulting genetic changes can impair the normal function of mitochondria, leading to the digestive and neurological problems associated with MNGIE. In some embodiments, treatment according to the present invention results in decreased intracellular and/or extracellular levels of thymidine in a patient by more than about 5%, e.g., about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, and about 100%, as compared to the average level of thymidine in the patient before the treatment or of one or more control individuals with similar disease without treatment. In various embodiments of the present disclosure, the method of reducing extracellular and intracellular levels of thymidine in a patient in need thereof, comprises administering a compound disclosed herein is effective for treating MNGIE.

[218] The terms, “improve,” “increase,” “reduce,” “decrease,” and the like, as used herein, indicate values that are relative to a control. In some embodiments, a suitable control is a baseline measurement, such as a measurement in the same individual prior to initiation of the treatment described herein, or a measurement in a control individual (or multiple control individuals) in the absence of the treatment described herein. A “control individual” is an

individual afflicted with MNGIE, who is about the same age and/or gender as the individual being treated (to ensure that the stages of the disease in the treated individual and the control individual(s) are comparable).

[219] The individual (also referred to as “patient”) being treated is an individual (fetus, infant, child, adolescent, or adult human) having MNGIE or having the potential to develop MNGIE. The individual can have residual endogenous thymidine phosphorylase expression and/or activity, or no measurable activity. In various embodiments, the individual having MNGIE may have thymidine phosphorylase expression or activity levels that are less than about 1-99% of normal thymidine phosphorylase expression or activity levels in an individual not afflicted with MNGIE. In some embodiments, the range includes, but is not limited to less than about 80-99%, less than about 65-80%, less than about 50-65%, less than about 30-50%, less than about 25-30%, less than about 20-25%, less than about 15-20%, less than about 10-15%, less than about 5-10%, less than about 1-5% of normal thymidine phosphorylase expression or activity levels.

[220] In some embodiments, the individual is an individual who has been recently diagnosed with the disease. Typically, early treatment (treatment commencing as soon as possible after diagnosis) is important to minimize the effects of the disease and to maximize the benefits of treatment.

Methods of Making

[221] The compounds described herein can be prepared in a variety of ways known to one skilled in the art of organic synthesis or variations thereon as appreciated by those skilled in the art. The compounds described herein can be prepared from readily available starting materials. Optimum reaction conditions can vary with the particular reactants or solvents used, but such conditions can be determined by one skilled in the art.

[222] Variations on the compounds described herein include the addition, subtraction, or movement of the various constituents as described for each compound. Similarly, when one or more chiral centers are present in a molecule, the chirality of the molecule can be changed. Additionally, compound synthesis can involve the protection and deprotection of various chemical groups. The use of protection and deprotection, and the selection of appropriate protecting groups can be determined by one skilled in the art. The chemistry of protecting groups can be found, for example, in Wuts and Greene, *Protective Groups in Organic*

Synthesis, 4th Ed., Wiley & Sons, 2006, which is incorporated herein by reference in its entirety.

[223] The starting materials and reagents used in preparing the disclosed compounds and compositions are either available from commercial suppliers such as Aldrich Chemical Co., (Milwaukee, WI), Acros Organics (Morris Plains, NJ), Fisher Scientific (Pittsburgh, PA), Sigma (St. Louis, MO), Pfizer (New York, NY), GlaxoSmithKline (Raleigh, NC), Merck (Whitehouse Station, NJ), Johnson & Johnson (New Brunswick, NJ), Aventis (Bridgewater, NJ), AstraZeneca (Wilmington, DE), Novartis (Basel, Switzerland), Wyeth (Madison, NJ), Bristol-Myers-Squibb (New York, NY), Roche (Basel, Switzerland), Lilly (Indianapolis, IN), Abbott (Abbott Park, IL), Schering Plough (Kenilworth, NJ), or Boehringer Ingelheim (Ingelheim, Germany), or are prepared by methods known to those skilled in the art following procedures set forth in references such as Fieser and Fieser's Reagents for Organic Synthesis, Volumes 1-17 (John Wiley and Sons, 1991); Rodd's Chemistry of Carbon Compounds, Volumes 1-5 and Supplementals (Elsevier Science Publishers, 1989); Organic Reactions, Volumes 1-40 (John Wiley and Sons, 1991); March's Advanced Organic Chemistry, (John Wiley and Sons, 4th Edition); and Larock's Comprehensive Organic Transformations (VCH Publishers Inc., 1989). Other materials, such as the pharmaceutical carriers disclosed herein can be obtained from commercial sources.

[224] Reactions to produce the compounds described herein can be carried out in solvents, which can be selected by one of skill in the art of organic synthesis. Solvents can be substantially nonreactive with the starting materials (reactants), the intermediates, or products under the conditions at which the reactions are carried out, *i.e.*, temperature and pressure. Reactions can be carried out in one solvent or a mixture of more than one solvent. Product or intermediate formation can be monitored according to any suitable method known in the art. For example, product formation can be monitored by spectroscopic means, such as nuclear magnetic resonance spectroscopy (*e.g.*, ^1H or ^{13}C) infrared spectroscopy, spectrophotometry (*e.g.*, UV-visible), or mass spectrometry, or by chromatography such as high performance liquid chromatography (HPLC) or thin layer chromatography.

[225] The disclosed compounds can be prepared by solid phase peptide synthesis wherein the amino acid α -N-terminal is protected by an acid or base protecting group. Such protecting groups should have the properties of being stable to the conditions of peptide linkage formation while being readily removable without destruction of the growing peptide chain or racemization of any of the chiral centers contained therein. Suitable protecting groups are 9-

fluorenylmethyloxycarbonyl (Fmoc), t-butyloxycarbonyl (Boc), benzyloxycarbonyl (Cbz), biphenylisopropylloxycarbonyl, t-amylloxycarbonyl, isobornyloxycarbonyl, α,α -dimethyl-3,5-dimethoxybenzyloxycarbonyl, o-nitrophenylsulfenyl, 2-cyano-t-butyloxycarbonyl, and the like. The 9-fluorenylmethyloxycarbonyl (Fmoc) protecting group is particularly preferred for the synthesis of the disclosed compounds. Other preferred side chain protecting groups are, for side chain amino groups like lysine and arginine, 2,2,5,7,8-pentamethylchroman-6-sulfonyl (pmc), nitro, p-toluenesulfonyl, 4-methoxybenzene-sulfonyl, Cbz, Boc, and adamantyloxycarbonyl; for tyrosine, benzyl, o-bromobenzyloxy-carbonyl, 2,6-dichlorobenzyl, isopropyl, t-butyl (t-Bu), cyclohexyl, cyclopentyl and acetyl (Ac); for serine, t-butyl, benzyl and tetrahydropyranyl; for histidine, trityl, benzyl, Cbz, p-toluenesulfonyl and 2,4-dinitrophenyl; for tryptophan, formyl; for aspartic acid and glutamic acid, benzyl and t-butyl and for cysteine, triphenylmethyl (trityl). In the solid phase peptide synthesis method, the α -C-terminal amino acid is attached to a suitable solid support or resin. Suitable solid supports useful for the above synthesis are those materials which are inert to the reagents and reaction conditions of the stepwise condensation-deprotection reactions, as well as being insoluble in the media used. Solid supports for synthesis of α -C-terminal carboxy peptides is 4-hydroxymethylphenoxymethyl-copoly(styrene-1% divinylbenzene) or 4-(2',4'-dimethoxyphenyl-Fmoc-aminomethyl)phenoxyacetamidoethyl resin available from Applied Biosystems (Foster City, Calif.). The α -C-terminal amino acid is coupled to the resin by means of N,N'-dicyclohexylcarbodiimide (DCC), N,N'-diisopropylcarbodiimide (DIC) or O-benzotriazol-1-yl-N,N,N',N'-tetramethyluroniumhexafluorophosphate (HBTU), with or without 4-dimethylaminopyridine (DMAP), 1-hydroxybenzotriazole (HOBT), benzotriazol-1-yloxy-tris(dimethylamino)phosphoniumhexafluorophosphate (BOP) or bis(2-oxo-3-oxazolidinyl)phosphine chloride (BOPCl), mediated coupling for from about 1 to about 24 hours at a temperature of between 10°C and 50°C in a solvent such as dichloromethane or DMF. When the solid support is 4-(2',4'-dimethoxyphenyl-Fmoc-aminomethyl)phenoxyacetamidoethyl resin, the Fmoc group is cleaved with a secondary amine, preferably piperidine, prior to coupling with the α -C-terminal amino acid as described above. One method for coupling to the deprotected 4-(2',4'-dimethoxyphenyl-Fmoc-aminomethyl)phenoxyacetamidoethyl resin is O-benzotriazol-1-yl-N,N,N',N'-tetramethyluroniumhexafluorophosphate (HBTU, 1 equiv.) and 1-hydroxybenzotriazole (HOBT, 1 equiv.) in DMF. The coupling of successive protected amino acids can be carried out in an automatic polypeptide synthesizer. In one example, the α -N-terminal in the amino

acids of the growing peptide chain are protected with Fmoc. The removal of the Fmoc protecting group from the α -N-terminal side of the growing peptide is accomplished by treatment with a secondary amine, preferably piperidine. Each protected amino acid is then introduced in about 3-fold molar excess, and the coupling is preferably carried out in DMF. The coupling agent can be O-benzotriazol-1-yl-N,N,N',N'-tetramethyluroniumhexafluorophosphate (HBTU, 1 equiv.) and 1-hydroxybenzotriazole (HOBT, 1 equiv.). At the end of the solid phase synthesis, the polypeptide is removed from the resin and deprotected, either successively or in a single operation. Removal of the polypeptide and deprotection can be accomplished in a single operation by treating the resin-bound polypeptide with a cleavage reagent comprising thioanisole, water, ethanedithiol and trifluoroacetic acid. In cases wherein the α -C-terminal of the polypeptide is an alkylamide, the resin is cleaved by aminolysis with an alkylamine. Alternatively, the peptide can be removed by transesterification, *e.g.* with methanol, followed by aminolysis or by direct transamidation. The protected peptide can be purified at this point or taken to the next step directly. The removal of the side chain protecting groups can be accomplished using the cleavage cocktail described above. The fully deprotected peptide can be purified by a sequence of chromatographic steps employing any or all of the following types: ion exchange on a weakly basic resin (acetate form); hydrophobic adsorption chromatography on underivatized polystyrene-divinylbenzene (for example, Amberlite XAD); silica gel adsorption chromatography; ion exchange chromatography on carboxymethylcellulose; partition chromatography, *e.g.* on Sephadex G-25, LH-20 or countercurrent distribution; high performance liquid chromatography (HPLC), especially reverse-phase HPLC on octyl- or octadecylsilyl-silica bonded phase column packing.

[226] The above polymers, such as PEG groups, can be attached to the TP under any suitable conditions used to react a protein with an activated polymer molecule. Any means known in the art can be used, including via acylation, reductive alkylation, Michael addition, thiol alkylation or other chemoselective conjugation/ligation methods through a reactive group on the PEG moiety (*e.g.*, an aldehyde, amino, ester, thiol, α -haloacetyl, maleimido or hydrazino group) to a reactive group on the TP (*e.g.*, an aldehyde, amino, ester, thiol, α -haloacetyl, maleimido or hydrazino group). Activating groups which can be used to link the water soluble polymer to one or more proteins include without limitation sulfone, maleimide, sulfhydryl, thiol, triflate, tresylate, aziridine, oxirane, 5-pyridyl, and α -halogenated acyl group (*e.g.*, α -iodo acetic acid, α -bromoacetic acid, α -chloroacetic acid). If attached to the TP by reductive

alkylation, the polymer selected should have a single reactive aldehyde so that the degree of polymerization is controlled. See, for example, Kinstler et al., *Adv. Drug. Delivery Rev.* 54: 477-485 (2002); Roberts et al., *Adv. Drug Delivery Rev.* 54: 459-476 (2002); and Zalipsky et al., *Adv. Drug Delivery Rev.* 16: 157-182 (1995).

[227] In order direct covalently link the TP to the CPP, appropriate amino acid residues of CPP may be reacted with an organic derivatizing agent that is capable of reacting with a selected side chain or the N- or C-termini of an amino acids. Reactive groups on the peptide or conjugate moiety include, e.g., an aldehyde, amino, ester, thiol, α -haloacetyl, maleimido or hydrazino group. Derivatizing agents include, for example, maleimidobenzoyl sulfosuccinimide ester (conjugation through cysteine residues), N-hydroxysuccinimide (through lysine residues), glutaraldehyde, succinic anhydride or other agents known in the art.

[228] The present disclosure also provides for recombinant fusion protein wherein a linear CPP is fused to the N-terminus and/or C-terminus of the TP. When prepared as recombinant fusions, the compounds can be prepared by known recombinant expression techniques. For example, to recombinantly produce the compound, a nucleic acid sequence encoding the chimeric gene is operatively linked to a suitable promoter sequence such that the nucleic acid sequence encoding such fusion protein will be transcribed and/or translated into the desired fusion protein in the host cells. Preferred promoters are those useful for expression in *E. coli*, such as the T7 promoter. Any commonly used expression system may be used, including eukaryotic or prokaryotic systems. Specific examples include yeast (e.g., *Saccharomyces spp.*, *Pichia spp.*), baculovirus, mammalian, and bacterial systems, such as *E. coli*, and *Caulobacter*.

Methods of Administration

[229] *In vivo* application of the disclosed compounds, and compositions containing them, can be accomplished by any suitable method and technique presently or prospectively known to those skilled in the art. For example, the disclosed compounds can be formulated in a physiologically- or pharmaceutically-acceptable form and administered by any suitable route known in the art including, for example, oral and parenteral routes of administration. As used herein, the term parenteral includes subcutaneous, intradermal, intravenous, intramuscular, intraperitoneal, and intrasternal administration, such as by injection. Administration of the disclosed compounds or compositions can be a single administration, or at continuous or distinct intervals as can be readily determined by a person skilled in the art.

[230] The compounds disclosed herein, and compositions comprising them, can also be administered utilizing liposome technology, slow release capsules, implantable pumps, and

biodegradable containers. These delivery methods can, advantageously, provide a uniform dosage over an extended period of time. The compounds can also be administered in their salt derivative forms or crystalline forms.

[231] The compounds disclosed herein can be formulated according to known methods for preparing pharmaceutically acceptable compositions. Formulations are described in detail in a number of sources which are well known and readily available to those skilled in the art. For example, *Remington's Pharmaceutical Science* by E.W. Martin (1995) describes formulations that can be used in connection with the disclosed methods. In general, the compounds disclosed herein can be formulated such that an effective amount of the compound is combined with a suitable carrier in order to facilitate effective administration of the compound. The compositions used can also be in a variety of forms. These include, for example, solid, semi-solid, and liquid dosage forms, such as tablets, pills, powders, liquid solutions or suspension, suppositories, injectable and infusible solutions, and sprays. The preferred form depends on the intended mode of administration and therapeutic application. The compositions also preferably include conventional pharmaceutically-acceptable carriers and diluents which are known to those skilled in the art. Examples of carriers or diluents for use with the compounds include ethanol, dimethyl sulfoxide, glycerol, alumina, starch, saline, and equivalent carriers and diluents. To provide for the administration of such dosages for the desired therapeutic treatment, compositions disclosed herein can advantageously comprise between about 0.1% and 100% by weight of the total of one or more of the subject compounds based on the weight of the total composition including carrier or diluent.

[232] Formulations suitable for administration include, for example, aqueous sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient; and aqueous and nonaqueous sterile suspensions, which can include suspending agents and thickening agents. The formulations can be presented in unit-dose or multi-dose containers, for example sealed ampoules and vials, and can be stored in a freeze dried (lyophilized) condition requiring only the condition of the sterile liquid carrier, for example, water for injections, prior to use. Extemporaneous injection solutions and suspensions can be prepared from sterile powder, granules, tablets, *etc.* It should be understood that in addition to the ingredients particularly mentioned above, the compositions disclosed herein can include other agents conventional in the art having regard to the type of formulation in question.

[233] Compounds disclosed herein, and compositions comprising them, can be delivered to a cell either through direct contact with the cell or via a carrier means. Carrier means for delivering compounds and compositions to cells are known in the art and include, for example, encapsulating the composition in a liposome moiety. Another means for delivery of compounds and compositions disclosed herein to a cell comprises attaching the compounds to a protein or nucleic acid that is targeted for delivery to the target cell. U.S. Patent No. 6,960,648 and U.S. Application Publication Nos. 20030032594 and 20020120100 disclose amino acid sequences that can be coupled to another composition and that allows the composition to be translocated across biological membranes. U.S. Application Publication No. 20020035243 also describes compositions for transporting biological moieties across cell membranes for intracellular delivery. Compounds can also be incorporated into polymers, examples of which include poly (D-L lactide-co-glycolide) polymer for intracranial tumors; poly[bis(p-carboxyphenoxy) propane:sebacic acid] in a 20:80 molar ratio (as used in GLIADEL); chondroitin; chitin; and chitosan.

[234] Compounds and compositions disclosed herein, including pharmaceutically acceptable salts or prodrugs thereof, can be administered intravenously, intramuscularly, or intraperitoneally by infusion or injection. Solutions of the active agent or its salts can be prepared in water, optionally mixed with a nontoxic surfactant. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, triacetin, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations can contain a preservative to prevent the growth of microorganisms.

[235] The pharmaceutical dosage forms suitable for injection or infusion can include sterile aqueous solutions or dispersions or sterile powders comprising the active ingredient, which are adapted for the extemporaneous preparation of sterile injectable or infusible solutions or dispersions, optionally encapsulated in liposomes. The ultimate dosage form should be sterile, fluid and stable under the conditions of manufacture and storage. The liquid carrier or vehicle can be a solvent or liquid dispersion medium comprising, for example, water, ethanol, a polyol (for example, glycerol, propylene glycol, liquid polyethylene glycols, and the like), vegetable oils, nontoxic glyceryl esters, and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the formation of liposomes, by the maintenance of the required particle size in the case of dispersions or by the use of surfactants. Optionally, the prevention of the action of microorganisms can be brought about by various other antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and

the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, buffers or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the inclusion of agents that delay absorption, for example, aluminum monostearate and gelatin.

[236] Sterile injectable solutions are prepared by incorporating a compound and/or agent disclosed herein in the required amount in the appropriate solvent with various other ingredients enumerated above, as required, followed by filter sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze drying techniques, which yield a powder of the active ingredient plus any additional desired ingredient present in the previously sterile-filtered solutions.

[237] Useful dosages of the compounds and agents and pharmaceutical compositions disclosed herein can be determined by comparing their *in vitro* activity, and *in vivo* activity in animal models. Methods for the extrapolation of effective dosages in mice, and other animals, to humans are known to the art.

[238] The dosage ranges for the administration of the compositions are those large enough to produce the desired effect in which the symptoms or disorder are affected. The dosage should not be so large as to cause adverse side effects, such as unwanted cross-reactions, anaphylactic reactions, and the like. Generally, the dosage will vary with the age, condition, sex and extent of the disease in the patient and can be determined by one of skill in the art. The dosage can be adjusted by the individual physician in the event of any counterindications. Dosage can vary, and can be administered in one or more dose administrations daily, for one or several days.

[239] Also disclosed are pharmaceutical compositions that comprise a compound disclosed herein in combination with a pharmaceutically acceptable carrier. Pharmaceutical compositions adapted for oral, topical or parenteral administration, comprising an amount of a compound constitute a preferred aspect. The dose administered to a patient, particularly a human, should be sufficient to achieve a therapeutic response in the patient over a reasonable time frame, without lethal toxicity, and preferably causing no more than an acceptable level of side effects or morbidity. One skilled in the art will recognize that dosage will depend upon a variety of factors including the condition (health) of the subject, the body weight of the subject, kind of concurrent treatment, if any, frequency of treatment, therapeutic ratio, as well as the severity and stage of the pathological condition.

[240] Also disclosed are kits that comprise a compound disclosed herein in one or more containers. The disclosed kits can optionally include pharmaceutically acceptable carriers and/or diluents. In one embodiment, a kit includes one or more other components, adjuncts, or adjuvants as described herein. In another embodiment, a kit includes one or more anti-cancer agents, such as those agents described herein. In one embodiment, a kit includes instructions or packaging materials that describe how to administer a compound or composition of the kit. Containers of the kit can be of any suitable material, *e.g.*, glass, plastic, metal, *etc.*, and of any suitable size, shape, or configuration. In one embodiment, a compound and/or agent disclosed herein is provided in the kit as a solid, such as a tablet, pill, or powder form. In another embodiment, a compound and/or agent disclosed herein is provided in the kit as a liquid or solution. In one embodiment, the kit comprises an ampoule or syringe containing a compound and/or agent disclosed herein in liquid or solution form.

[241] A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

EXAMPLES

Example 1. Synthesis of Compounds

[242] **Thymidine phosphorylase.** The gene coding for the mature thymidine phosphorylase protein (11-482) was prepared by de novo gene synthesis, and the resulting DNA fragment was subcloned in a prokaryotic expression vector pET-30a(+) at EcoRV-EcoRI sites. *E. coli* Lemo21(DE3) competent cells transformed with the plasmid encoding the thymidine phosphorylase was incubated at 37 °C in a LB containing 50 µg/mL kanamycin. The culture was grown at 37 °C until an OD600 between 0.4-0.6. The protein expression was induced at 25 °C overnight in the presence of 0.25 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). After overnight culture, bacteria cells were harvested by centrifugation (4000 g for 15 min at 4 °C). Cell pellets were stored at -20 °C until further purification.

[243] The gene coding for truncated active thymidine phosphorylase protein (TP16, TP22, or TP35) were prepared by de novo gene synthesis, and the resulting DNA fragment was subcloned in a prokaryotic expression vector pET-24a(+) at NdeI-XhoI sites. EMD Millipore™ Novagen™ Rosetta™ 2 (DE3) Singles Competent Cells transformed with the plasmid encoding the thymidine phosphorylase was incubated at 37 °C in a LB containing 50 µg/mL kanamycin in the shaking flask. The culture was grown at 37 °C until an OD600 between 0.4-0.6. The protein expression was induced at 30 °C overnight in the presence of 0.25 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). After overnight culture, bacteria cells were harvested by centrifugation (4000 g for 15 min at 4 °C). Cell pellets were stored at -20 °C until further purification.

[244] To prepare human IgG1-Fc fusion of TP proteins, the gene coding for human IgG1Fc fused TP conjugates (SEQ ID NO. 6-13, Table 6) were prepared by de novo gene synthesis, and the resulting DNA fragment was subcloned in a prokaryotic expression vector pET-21a(+) at NdeI-XhoI sites. EMD Millipore™ Novagen™ Rosetta™ 2 (DE3) Singles Competent Cells transformed with the plasmid encoding SEQ ID NO 7 (Fc-TP16) was incubated at 37 °C in a LB containing 50 µg/mL Ampicillin in the shaking flask. The culture was grown at 37 °C until an OD600 between 0.4-0.6. The protein expression was induced at 30 °C overnight in the presence of 0.25 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). After overnight culture, bacteria cells were harvested by centrifugation (4000 g for 15 min at 4 °C). Cell pellets were stored at -20 °C until further purification.

[245] Process of preparing recombinant TP protein using bioreactor: a glycerol stock seed (1 mL) was thawed and used to inoculate the initial culture (5 ml growth medium composed of 10 g/L soyton, 5 g/L yeast extract, 10 g/L NaCl, pH 7.5, 50 mg/L Kanamycin) at 37°C, 250 rpm for 4 hour. Then the culture was used to inoculate 100 mL medium and grown additional 5 hours. When the culture achieved an optical density (OD_{600nm}) around 2-3, it was used to inoculate a 2 L minimum medium in the bioreactor. The composition of the minimum medium per liter is as follows: 2 g (NH₄)₂SO₄, 6.75 g KH₂PO₄, 0.35 g MgSO₄, 0.85 g citric acid, 20 g/L glucose. After inoculation, the culture was grown for 10 hours at 37°C with the temperature controlled by a PID loop. Dissolved oxygen (DO) was set at 30% and was also controlled by a PID loop control and with stirring-first-oxygen priority in that order, the minimum/maximum stirring was set at 400/800 rpm. The pH was adjusted to 7.2 and controlled by a PID loop, using 28-30% (m/v) NH₄OH through the alkali pump. After 10 hours, the OD_{600nm} reached about 9-10, a glucose feeding solution was set up through the acid pump and under the same pH PID

loop control, and this was maintained until the end of the fermentation process. The composition of the glucose feeding solution is as follows: 50% (w/v) glucose, 9.8 g/L $(\text{NH}_4)_2\text{SO}_4$ and 1x trace metal solution. Additionally, a yeast extract solution was constantly fed through an external pump at 10-15 mL per hour per liter of culture. The composition of this solution was 10 g/L yeast extract and 20 g/L NaH_2PO_4 . The culture was induced with 0.5 mM IPTG at $\text{OD}_{600\text{nm}}$ around 20 at 30 degree Celsius. The culture was induced two more times with the same IPTG concentration every 2.5 hours. The biomass was harvested 8-9 hours post-first-induction by centrifugation, the achieved biomass production was ~50 g/L of culture, and the pellet was stored at -80°C till further purification.

[246] For the purification of TP11, bacteria pellets from 1 L cell culture were resuspended in 50 mL lysis buffer (20 mM sodium phosphate, pH 7.4, 200 mM sodium chloride, 5% (v/v) glycerol, 20 mM imidazole). Once the pellet was resuspended, 5 mg of pre-dissolved lysozyme and 500 μL of protease inhibitor cocktail (protease inhibitor cocktail, Sigma, P8849) were added. The solution was stirred on ice for approximately 15 min and the cells were sonicated on ice (pulsed six times for 20 sec with 40 sec resting periods in between, level 100; Sonic Dismembrator, Model 100, Fisher Scientific). The suspension was cleared by centrifugation at 20,000 g for 30 min at 4°C . The supernatant was harvested and applied to a 5 ml His-tag affinity column (HisTrap, fast flow, 5 ml, GE) at 1 mL/min before washed with 100 mL of washing buffer A (20 mM sodium phosphate, pH 7.4, 200 mM sodium chloride, 5% (v/v) glycerol, 20 mM imidazole). To remove endotoxin, the column was then washed with 20 column volume of Triton X-114 buffer (buffer A supplemented with 0.1% (v/v) Triton X-114 detergent) followed by 20 column volume of CHAPS buffer (buffer A supplemented with 1% (w/v) CHAPS detergent). Afterwards, the column was washed with 10 CV of buffer A before His-tagged thymidine phosphorylase was eluted with buffer A supplemented with 100-500 mM imidazole in 10 column volume (FIG. 1A).

[247] The purity of TP was determined by SDS-PAGE and fractions of high purity were combined. Endotoxin level was typically lower than 50 EU per mg quantified by the Pierce™ LAL Chromogenic Endotoxin Quantitation Kit. Purified His-tagged TP11 were pooled and dialyzed against 4 liter of EK cleavage buffer (20 mM Tris-HCl, pH 7.4, 50 mM NaCl, 2 mM CaCl_2 , 5% glycerol). To remove the affinity tag, 100 mg of His-tagged protein at a concentration of 2 mg/mL was mixed with 1:100 (w/w) recombinant bovine Enterokinase (His-tagged) (GeneScript) at 4°C overnight. The cleavage mixture were then loaded to 5 mL HisTrap column. The flow-through and 5 CV of washing solution typically contain tag-free

thymidine phosphorylase, which were dialyzed against phosphate buffer saline for the bioconjugation step (FIG. 1B).

[248] For the purification of tag free TP16 harvested from bioreactor process described above. The pellet in lysis buffer (5 mL per gram of biomass). The composition of the lysis buffer is 50 mM Tris-HCl pH 7.5, 5 mM EDTA, 1 ml of 1X protease inhibitory cocktail per 20 grams of pellet. Homogenization of the biomass was performed by sonication in a Branson instrument, four rounds of 10-cycles sonication were required to fully release soluble TP enzyme. Sonication cycles were 30s On/30s Off at 30% intensity. Clarification of the crude solution was done by centrifugation for 30 min at 18000G and 4°C. Purification of TP16 from the crude supernatant was achieved in two chromatography steps using phenyl (hs) and capto adhere prepacked resins. First, the crude supernatant containing the soluble TP16 was loaded on a HiTrap® Phenyl Fast Flow (HS) in the presence of 0.7 M $(\text{NH}_4)_2\text{SO}_4$ with a proportion of 20 mg total protein per ml of resin. The phenyl column was pre-equilibrated with buffer A, 50 mM Tris-HCl 0.7 M $(\text{NH}_4)_2\text{SO}_4$ adjusted to pH 7.5. The bound protein was washed with 10 column volume (CV) of buffer A. The target enzyme was eluted with a three-steps gradient (50%, 75% and 100%) of increasing concentration of buffer B, 10 mM TrisHCl pH 8.0 (FIG. 17A). The fraction that eluted at 75% of buffer B was directly loaded into capto adhere prepacked column with a proportion of 10 mg/ml of resin. The capto adhere column was pre-equilibrated with buffer B. Bound protein was washed with 20 CV of buffer B, followed by 20 CV of 50 mM citrate buffer washing, then 1 M NaCl pH 4.5, 20 CV of PBS buffer washing, 0.250 M arginine pH 7.4 washing, and 20 CV of 0.5% CHAPS 5 mM EDTA in TrisHCl buffer pH 7.5 washing. Most of the target protein was eluted with 30 CV of 0.500 M arginine in PBS buffer pH 7.4, while the rest was eluted with 20 CV of 0.75 M guanidine-HCl. The purity of this pool was more than 90% (determined by SDS-PAGE) (FIG. 17B). The final TP16 production was buffer exchanged into PBS for storage or protein modifications.

[249] The tag free TP can also be produced with high efficiency using an inclusion body based refolding process. In this case, a glycerol stock seed (1 mL) was thawed and used to inoculate the initial culture (5 ml terrific broth growth medium, composed of 10 g/L soyton, 5 g/L yeast extract, 10 g/L NaCl, pH 7.5, 50 mg/Ll Kanamycin) at 28°C, 250 rpm for overnight. When the culture achieved an optical density ($\text{OD}_{600\text{nm}}$) around 2-3, it was used to inoculate a 2 L terrific broth growth medium in the bioreactor. After inoculation, the culture was grown at 37°C with the temperature controlled by a PID loop. Dissolved oxygen (DO) was set at 20% and was also controlled by a PID loop control and with stirring-first-oxygen priority in that

order, the minimum/maximum stirring was set at 400/800 rpm. The pH was adjusted to 7.2 and controlled by a PID loop, using 28-30% (m/v) NH_4OH through the alkali pump. A glucose feeding solution was set up through the acid pump and under the same pH PID loop control, and this was maintained until the end of the fermentation process. The culture was induced with 1 mM IPTG at $\text{OD}_{600\text{nm}}$ around 15 at 37 degree Celsius. The culture was induced one more time with the same IPTG concentration after 2.5 hours. The biomass was harvested 5-6 hours post-first-induction by centrifugation, the achieved biomass production was ~25 g/L of culture, and the pellet was stored at -80 °C till further purification.

[250] The cell pellet was resuspended in 50 mM Tris buffer pH 8.0, 200 mM NaCl, 5 mM DTT and homogenized by sonication (10 mL lysis buffer per gram pellet). The lysate was centrifuged at 12000 rpm for 30 min and the pellet containing inclusion bodies were collected. The inclusion bodies were washed with a buffer containing 50mM Tris, pH 8.0, 200 mM NaCl, 5mM DTT and 1% Triton X-100, resuspended by a short cycle of sonication and centrifuged at 12000 rpm at 4 °C. Four washing steps were needed to remove most of impure proteins and membrane components. A buffer without Triton but with 1M NaCl was used for the last step washing to remove residual Triton X-100 and host genomic DNA. The purified inclusion bodies can be stored at -20 °C till further purification.

[251] The inclusion body was gently dissolved with 50 mM Tris, pH 8.0, 200 mM NaCl, 8 M Urea, and 10 mM DTT to 20 mg/mL of protein solution at room temperature. The dissolved inclusion body solution was harvested as the soluble solution after centrifuge at 12000 rpm for 30 min at 4 °C. Afterwards, urea-free buffer was added to bring down the urea concentration to 6M. To remove the bioburden and DNA contents, the inclusion body solution was applied on Q sepharose column on an AKTA purifier and the flow through was collected. To refold the protein, the inclusion body solution was rapidly diluted by 20 fold into 50 mM Tris, pH 8.0, 200 mM NaCl, and 5 mM DTT at 4 °C. The resulting solution was kept at 4 °C for overnight before it was further diluted another six fold with into 50 mM Tris, pH 8.0 and 1 mM DTT. The resulting solution was loaded on Q sepharase column and the target protein was eluted with increasing sodium chloride concentration to obtain the refolded protein with desired specific TP activity (FIG. 35 and 36).

[252] For the purification of human IgG1Fc fusion TP16 (e.g. Fc-TP16), bacteria pellets from 1 L cell culture were resuspended in 10 mL lysis buffer (20 mM sodium phosphate, pH 7.4, 1mM EDTA). Once the pellet was resuspended, 2 mg of pre-dissolved lysozyme and 100 μL of protease inhibitor cocktail (protease inhibitor cocktail, Sigma, P8849) were added. The solution

was stirred on ice for approximately 15 min and the cells were sonicated on ice. The suspension was cleared by centrifugation at 20,000 g for 30 min at 4 °C. The supernatant was harvested and applied to a 1 ml Protein A column (Protein A, fast flow, 1 ml, GE) at 1 mL/min before washed with 20 mL of washing buffer A (20 mM sodium phosphate, pH 7.4). To remove endotoxin, the column was then washed with 50 column volume of Triton X-114 buffer and CHAPS buffer (buffer A supplemented with 0.1% Triton X-114 (v/v) and 1% (w/v) CHAPS detergent). Afterwards, the column was washed with 20 CV of buffer A before elution with elution buffer (0.1M sodium citrate, pH 3.0) in 10 column volume (FIG. 30).

[253] Cyclic cell penetrating peptides (cCPPs). cCPPs were synthesized by solid phase peptide synthesis using Fmoc-chemistry, deprotected and released from the solid support, triturated, and purified using RP-HPLC. Conjugation of between cCPP12 with a C-terminal Lysine and 4-formyl-benzamido-dPEG₁₂-TFP ester (Product # 10081, Quanta Biodesign) was performed in pH 7.4 phosphate buffer at 1:1 ratio for 2 h. The product, cCPP12-PEG₁₂-FBA, was again purified by RP-HPLC and lyophilized for storage prior to use.

[254] Amine-based TP-cCPP conjugation. To prepare the cCPP-TP conjugates through reductive amination reactions on amine groups of TP protein, freshly purified TP11 or TP16 (0.5 mg/mL, 10 µM) was mixed with the cCPP-linker conjugate, e.g., cCPP12-PEG₁₂-FBA (80 µM) in pH 6.0 2-(N-morpholino)ethanesulfonic acid buffer (0.1 M), followed by the addition of 10 mM freshly prepared sodium cyanoborohydride. The reaction was gently mixed for 36 h before analyzed by SDS-PAGE to confirm the completion of bioconjugation. Reaction was then quenched with glycine, and small molecules as well as extra peptides were removed by dialysis against phosphate buffer saline (pH 7.4) for 16 h twice. The resulting conjugates are represented as “CPP-N-TP”, with N referring to the N-terminal of TP as the site of conjugation. Alternatively, cCPP12-N-TP11 or cCPP12-N-TP16 were prepared by mixing TP (2 mg/mL, 40 µM) with cCPP12-PEG₁₂-FBA (320 µM) and 10 mM freshly prepared sodium cyanoborohydride in phosphate buffer saline (pH 7.4) for 36 h, and then the reaction mixture was purified by filtration to remove extra peptide and other chemical reagents. Conjugated protein (e.g. cCPP12-N-TP11, *see* FIG. 3 or FIG. 19; e.g. cCPP12-N-TP16, *see* FIG. 19) was then temporally stored in 4 °C for immediate usage or formulated in the presence of 2% mannitol for long-term storage at -20 °C.

[255] To prepare the CPP-conjugated Fc-TP16 through reductive amination reactions on amine groups of TP protein, freshly purified FcTP16 (1.65mg/mL, 22 micromolar) was mixed with the cCPP-linker conjugate, e.g., cCPP-PEG₁₂-FBA (220 micromolar) in pH 8.0, 4-(2-

hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (0.1 M), followed by the addition of 10 mM freshly prepared sodium cyanoborohydride. The reaction was gently mixed for 48 h before analyzed by SDS-PAGE to confirm the completion of bioconjugation (e.g. cCPP12-N-Fc-TP16, *see* FIG. 30).

[256] PEGylated Conjugates. To prepare PEGylated products, freshly purified cCPP12-N-TP11 or cCPP12-N-TP16 (4 mg/mL, 80 μ M) was mixed with the PEG5K/10K/40K linear/40K branched-NHS ester (NHS, N-hydroxyl succinimide) or PEG12-NHS ester (around 2 kDa molecular weight). The reaction was gently mixed for 2 h at room temperature in phosphate buffer (50 mM sodium phosphate, pH 7.4, 150 mM sodium chloride) or sodium bicarbonate buffer before analyzed by SDS-PAGE to confirm the completion of PEGylation. PEGylated proteins (CPP12-N-TP-PEG5K, CPP12-N-TP-PEG10K, CPP12-N-TP-PEG40K linear, CPP12-N-TP-PEG40K branched, or CPP12-N-TP-PEG12) were then diluted with 20 mM Tris, pH 8.0 to 0.4 mg/mL protein concentration before applied on Q-Sepharose column at the flow rate of 1 ml/min. Additional PEGylation reagents were washed with 20 mM Tris, pH 8.0. PEGylated proteins were eluted with 20 mM Tris, pH 8.0 with 1 M sodium chloride. Combined fractions with desired product were dialyzed with phosphate buffer saline (pH 7.4) twice and sterile filtered and stored at 4 mg/mL at -20 °C. Final products including cCPP12-N-TP-PEG5K, cCPP12-N-TP-PEG10K, cCPP12-N-TP-PEG40K linear, or cCPP12-N-TP-PEG40K branch were further characterized in biological assays.

[257] To facilitate the detection of TP and conjugated TP in vitro and in vivo. Alexa568 fluorophore was used to label the protein on random lysine residues. Briefly, 2.0 mg/mL of protein was mixed with 7.5-15 equivalent of Alexa Fluor 568 NHS Ester (ThermoFisher Scientific) at room temperature. The reaction was quenched by glycine solution after 2 h. Fluorescently labeled material was isolated by size exclusion chromatography.

[258] Disulfide-based TP-cCPP Conjugation (“CPP-S-S-TP”). To conjugate the TP protein through Cys residues, cyclic CPP with an activated disulfide modality (SPDP) was designed and synthesized on solid phase. The product, cCPP12-S-S-SPDP was purified by RP-HPLC and lyophilized for storage prior to conjugation. The structure is shown in FIG. 18.

[259] Preparation of cCPP12-SS-TP-PEG10K. Freshly purified TP11 or TP16 (2 mg/mL, 40 μ M) was mixed with the PEG10K-NHS ester (NHS, N-hydroxyl succinimide, 1:20 molar ratio) in PBS buffer. The reaction was gently mixed for 30 min at room temperature followed by another treatment of 1:20 molar ratio PEG10K-NHS ester for 1h at room temperature in PBS before analyzed by SDS-PAGE to confirm the completion of PEGylation. PEGylated

proteins (PEG10K-TP11 or PEG10K-TP16) were then diluted with PBS to 0.5 mg/mL protein concentration before cCPP conjugation. Diluted PEGylated proteins were pre-treated at 48 °C for 30min. cCPP12-B-B-SPDP was immediately added to the protein solution (1:30 molar ratio) and react at 48 °C for another 1h. Excess PEG10K and peptides were removed by Amicon Ultra-15 centrifugal filter units (30K MWCO). The conjugation reactions were monitored by RP-HPLC equipped with C4 column using 0.1% TFA (v/v) supplemented water as buffer A and 0.1% TFA (v/v) supplemented acetonitrile as buffer B as shown in FIG. 20. The products, CPP12-SS-TP11-PEG10K and CPP12-SS-TP16-PEG10K, were further evaluated in biological assays.

Example 2. Characterization of TP and TP conjugates

[260] A biochemical TP enzyme activity assay was developed. The enzyme activities of purified TP or conjugated TP were performed by measuring the change of absorbance at 290 nm during enzymatic phosphorylation of thymidine. Basically, 200 microliter of nanomolar concentrations of the enzyme in phosphate buffer saline was mixed with 2 mM thymidine at 37 °C. Progress of reactions were monitored at 290 nm, indicating the conversion of thymidine to thymine by TP. Coeffeciency of $2000 \text{ M}^{-1} \text{ cm}^{-1}$ was used to calculate the turnover rate in the unit of s^{-1} , which is the number of thymidine molecule converted to thymine every second by one enzyme. TP16 is functionally similar to TP11 at a concentration of 100 nM in the enzymatic assay (see FIG. 21, y axis represent the normalized enzyme activity, error bars represent the technical triplicates for the data).

[261] Enzyme stability of TP11 in serum was evaluated and the data is presented in FIG. 2B. The data in the graph indicates that enzymatic activity (i.e. enzymatic stability) of TP in mouse serum was maintained after both 2 h and 4 h incubation at 37 °C.

[262] CPP12 was conjugated to N-terminus of TP11 according to the procedures described in Example 1. The enzymatic activity of the resulting product CPP12-N-TP11 (see FIG. 3) was compared to unconjugated TP11 (FIG. 4A) at a concentration of 40 nM. Measuring enzyme turnover rate showed that CPP12-N-TP11 is functionally equivalent to unconjugated TP11 (similar observations for TP16 and CPP12-N-TP16). All PEGylated TP disulfide linked or N-terminus labeled CPP conjugates (CPP12-SS-TP11-PEG10K and CPP12-SS-TP16-PEG10K and CPP12-N-TP11-PEG10K and CPP12-N-TP16-PEG10K) possess similar activity to unconjugated TP11 at a concentration of 100 nM in enzymatic assay (FIG. 22). And Fc fused TP16 (e.g. Fc-TP16) and the CPP conjugated Fc-TP16 also show similar enzyme activity compared to unconjugated TP11 at a concentration of 100nM in enzymatic assay (FIG. 31).

[263] Further characterization of CPP12-N-TP11 showed that the conjugated compound is stable after freezing and thawing (FIG. 4B). It also remains enzymatically stable after incubation with full growth medium (DMEM +FBS) as well as mouse serum (FIG. 4C). In addition, no degradation was observed for either Alexa568-labeled TP11 or CPP12-N-TP11 in mouse serum at 37 °C for 2 h, 12 h, and 24 h (FIG. 5).

Example 3. Cellular Delivery of TP

[264] Studies showed that CPP12-N-TP efficiently enters cells and is enzymatically active. LS174T cells (1.5×10^6 per well) were seeded on a 6-well plate in McCoy 5A modified medium containing 10% FBS and 1% penicillin/streptomycin, and cultured for 16 h to reach approximately 80% confluency. Then the growth medium for each well was replaced by 1.5 mL of fresh medium with or without designated amount of TP or CPP12-N-TP. After 6 h treatment, the medium was removed and each well was washed by 4 ml phosphate buffer saline four times. The cells were harvested by trypsinization, and pelleted at 250 G for 5 min. Resuspended cells again were washed with phosphate buffer saline and pelleted for storage at -80 °C till further analysis. Different cell pellets were resuspended on ice with 100 microliter of lysis buffer to extract either cytosolic proteins using cytosolic lysis buffer (50 mg/mL digitonin, 75 mM sodium chloride, 10 mM sodium phosphate, pH 7.4, 250 mM sucrose supplemented with protease inhibitors) or whole-cell proteins using whole cell lysis buffer (1% Triton X-100, 150 mM sodium chloride, 50 mM Tris-HCl, pH 8.0). After lysis, cellular contents were then centrifuged at 16,000 G for 10 min. Supernatants were collected for western blotting analysis or enzyme activity analysis. Western Blotting analysis (FIG. 6A) was used to compare intracellular levels of TP in Hela Cells (WT) with intracellular levels of TP in LS174T (TP-deficient) cells treated under three sets of conditions. Lane 1 showed a high concentration of transduced TP inside of LS174T cells after treatment of 6 h with 1 μ M CPP12-N-TP, yielding more abundant TP protein than that of WT cells. Minimum amount of TP entered LS174T cells that were treated with 1 μ M unconjugated TP. This level is in line with what was established with control medium. The Western Blotting analysis demonstrated that the CPP modification is necessary and efficient to deliver TP into mammalian cells.

[265] Intracellular TP activity of cell lysates is expressed as the thymidine to thymine conversion (in nanomole) per hour per mg of cell lysate protein. Briefly, 10 microliter of 10X phosphate buffer saline, pH 7.4, and 12.5 microliter of 80 mM thymidine were added into 77.5 microliter of cell lysate proteins of approximately 2 mg/mL concentration. The reaction

mixtures were incubated at 37 °C for different time periods: 0 hr, 1 hr, 4 hr or overnight. After desired reaction time, 20 microliter of reaction mixture was mixed with 180 microliter of 0.3 M sodium hydroxide in water to terminate the reaction. The absorbance at 300 nm was measured, and thymidine to thymine conversion was calculated using absorption coefficient of 3,400 M⁻¹cm⁻¹. Comparison of intracellular TP activity (FIG. 6B) between Hela cells (WT) and TP deficient cells (LS174T) under various treatment demonstrated efficient and functional intracellular delivery of TP after 6 h treatment with 1 μM CPP12-N-TP. The enzyme activity of CPP12-N-TP treated LS174T cells is over 300 nmol hr⁻¹ mg⁻¹, which is considerably stronger than that of WT cells. TP treated and medium treated deficient cells only have background level of TP activity which is between 20-40 nmol hr⁻¹ mg⁻¹. Thus, CPP12-N-TP not only enters cells, but exhibits high levels of enzymatic activity.

[266] Dose dependent delivery of TP into LS174 cells was analyzed by Western Blot (FIG. 7). Modifying the amount of CPP12-N-TP used to incubate the cells from 0.1 to 0.5 to 1 μM CPP12-N-TP showed a dose-dependent increase in the level of TP delivered into the cultured cells. A substantial increase of intracellular TP contents s observed with 0.5 or 1 μM CPP12-N-TP compared to cells incubated with 1 μM TP, which was essentially equivalent to media treated cells (negative control).

[267] Whole cell and cytosolic TP activity delivered into TP-deficient LS174T cells were analyzed by a TP enzyme activity assay (FIG. 8). LS174T cells were treated for 6 h with different concentrations of CPP12-N-TP (0.1, 0.5, or 1 μM). 1 μM TP and medium treatment were used as controls. The whole cell lysate samples were collected and characterized by enzyme activity assay. Dose-dependency was verified by the data presented in columns 3, 4, and 5, which showed the intracellular TP activity of LS174T cells after treatment with 0.1, 0.5, or 1 μM of CPP12-N-TP, respectively. Cellular uptake and intracellular TP activity was highest at 1 μM CPP12-N-TP, as expected based on the WB results. Moreover, for 1 micromolar CPP12-N-TP treated cells, its cytosolic fraction was also collected. The cytosolic TP activity (column 6) was comparable to the TP activity from WT cells. Comparing TP activity from columns 5, and 6, more than 50% delivered enzyme activity entered cytosol. The result further confirmed that CPP12-N-TP not only successfully internalized into the cells, but also efficiently escaped from the endosome and entered into cytosol.

[268] To evaluate the uptake efficiency of various CPP-TP-conjugates, whole cell and cytosolic TP activity delivered into TP-deficient LS174T cells were analyzed by a similar TP enzyme activity assay. LS174T cells were treated for 6 h with 1 micromolar of cCPP12-N-

TP11, cCPP12-N-TP11-PEG10K, cCPP12-N-TP16-PEG10K, cCPP12-SS-TP11-PEG10K, or cCPP12-SS-TP16-PEG10K. 1 μ M TP and medium treatment were used as controls. The cytosolic fraction of incubated cells were collected and characterized by enzyme activity assay (FIG. 34). Cellular uptake and intracellular TP activity were highest at 1 μ M CPP12-N-TP treated cells. cCPP12-N-TP11-PEG10K as well as cCPP12-N-TP16-PEG10K also showed significant uptake as expected. Surprisingly, the disulfide based conjugated cCPP12-SS-TP11-PEG10K and cCPP12-SS-TP16-PEG10K showed minimal cellular uptake similar to the level of tag free TP. The result demonstrated that the choice of conjugation chemistry and the site of conjugation are critical for successful cellular delivery and endosomal escape into cytosol.

Example 4. In Vivo Assay

[269] To determine the half-life and biodistribution of CPP—conjugated TP proteins, Alexa568-labeled proteins will be injected through subcutaneous (s.c.), intradermal (i.d.), intravenous (i.v.), or intraperitoneal (i.p.) routes into CD1 mouse or C57BL/6 mouse at 0.1, 1, 2, 5, 10, or 20 mpk per injection. Control group will be injected with PBS or Alexa568-labeled TP protein. Plasma, blood cells, PBMC, various organs (heart, lung, liver, spleen, pancreas, kidney, muscle, intestine, and brain) will be harvested at various time points post injection (0.5, 4, 8, 24, 48, or 96 h). TP enzyme activities from various tissues samples were quantified. Tissue samples were properly reserved for further immunohistology analysis as well as histopathology analysis. Fluorescence in various tissues were quantified. Biodistribution of fluorescently-labeled TP proteins were further examined by SDS-PAGE analysis of tissue homogenates followed by in-gel fluorescence scanning.

[270] To demonstrate the in vivo efficacy of CPP-conjugated TP as a potential treatment of MNGIE, a murine model was leveraged: Tymp/Upp1 double KO mice (reference: Hum Mol Genet 18: 714–722.). Tymp/Upp1 double KO mice were injected with CPP-conjugated TP, TP, or solvent cohorts by i.p., i.v., or s.c. method at once, twice, or thrice weekly with dose up to 20 mpk per injection. Blood samples were collected weekly. Thymidine and deoxyuridine levels in blood were tested by LC-MS assay throughout the study. Thymidine phosphorylase activity in blood were also tested by the TP enzyme activity assay throughout the study. By the end of the treatment (4-12 weeks), mice were sacrificed and the thymidine and level in liver, brain, skeletal muscle, small intestine, and kidney were quantified by LC-MS assay. From the same tissue samples, TP enzyme activity was also quantified following literature method (ref. FEBS Lett 581: 3410–3414.). The abundance of delivered TP in various tissues was also analyzed by western blotting as well as by immunofluorescence assays.

[271] After administration of conjugated TP, plasma thymidine concentrations in most study samples drop from relatively high level (around 10 μ M in MNGIE mouse model) to wild-type level (1-3 micromolar). Two to eight months after treatment (end of the study), nucleotides levels in tissues harvested also drop significantly. For example, in liver, thymidine level will decrease from 40-120 pmol/mg protein to lower than 20 pmol/mg protein. Substantial increases of TP protein in various tissues are detected by both an enzyme activity assay as well as Western Blot analysis.

Example 5. Quantification of Serum Thymidine Level in MNGIE Mouse Model

[272] MNGIE mouse models (Mol Cell Biol. 2002; 22: 5212-5221) were used to evaluate thymidine reduction in serum by administrations of CPP-TP conjugates disclosed herein. MNGIE mice have aberrantly high levels (around 10 μ M) of thymidine due to the absence of functional *TYMP* and *UPP1* genes thus cannot metabolize thymidine into thymine effectively. Delivery of TP via CPP is therefore expected to reduce thymidine levels.

[273] To quantify the thymidine level in serum, approximately 25 microliter of freshly isolated serum sample were mixed with 46.8 microliter of distilled water and 3.2 microliter of concentrated perchloric acid (initial concentration 11.7 M, final concentration of 0.55 M). Samples were then vortexed for 10 s and kept on ice for 10 min to help with protein precipitation. Afterwards, precipitates were removed by centrifuge at 17,000 G at 4 °C for 10 min. Clear supernatants were collected and analyzed on an Agilent 1100 analytical HPLC equipped with a C18 5 μ m, 4.6X250 mm column using gradient elution and UV detection (268 nm). Elution gradient can be referenced to Methods Mol. Biol. 2012; 837: 121-133. Concentration of thymidine, deoxyuridine, and other nucleotides were calculated using area under curve (AUC) and the calibration curves.

[274] To test the ability of CPP-TP conjugates to reduce thymidine levels, MNGIE mice were treated with a CPP-TP conjugate via tail-vein intravenous injection. Treatment occurred on four days: day 0; day 5; day 12 and day 19 at 7.5, 7.5, 20, and 20 mpk respectively. Serum thymidine levels of the MNGIE mice were measured two days prior to treatment (day -2), and the day after each treatment (day 1, after the first injection; day 6, after the second injection; day 12, after the third injection; and day 20, after the fourth injection). Thymidine levels were compared to untreated MNGIE mice. Statistical analysis was performed using a one-way ANOVA with Geisser-Greenhouse correction (***) indicates a p-value < 0.0001). The results are provided in FIG. 9.

[275] The results show that the CPP-TP conjugates reduce serum thymidine levels to healthy levels measured for wild-type mice (comparing day-2 levels to day20 level) whereas thymidine levels in control (phosphate buffer saline) treated group remained elevated.

[276] To confirm results obtained on day 20, the same groups of MNGIE mice were injected again with 20 mpk CPP-TP or vehicle control on day 26, and their thymidine levels were measured on day 27 as described above. Interestingly, three out of four (other than mouse T20) treated MNGIE mice showed depleted thymidine levels in serum, which further confirmed the in vivo efficacy of CPP-TP treatments (FIG. 10A).

[277] To investigate the incomplete depletion of thymidine level in the serum of MNGIE mouse T20 (FIG. 10A), potential immunogenicity properties of CPP-human TP were investigated by detecting the formation of anti-drug antibody using an ELISA assay. Briefly, wells in 96-well polystyrene plate were coated with 1 µg/mL of CPP-TP at 37 °C for 1 hour. Wells were then washed three times with washing buffer (phosphate buffer saline and 0.05% tween 20) and incubated with blocking buffer (phosphate buffer saline and 2% bovine serum albumin) at 37 °C for 0.5 hour. Wells were then washed once with washing buffer before incubated with serially diluted serum samples from treated (T9, T10, T15, or T20) or control (C8, C11, or C16) mice at 37 °C for 1 hour. After the incubation, wells were washed thrice and then incubated with 1 µg/mL HRP-labeled goat-anti-mouse IgG (H+L) at 4 °C. Afterwards, wells were washed thrice and incubated with TMB substrate solution at room temperature in the dark for 30 min before quenched with 0.1 mL of 1 M hydrogen chloride solution. The absorption at 450 nm were recorded by a plate reader, and the values were plotted against the dilution factor of plasma (FIG. 10B). In this ELISA assay, stronger AFU (i.e. absorption at 450 nm) is positively correlated with increased anti-CPP-TP mouse IgG antibody. Interestingly, serum from mouse T20 is showed significantly elevated level of anti-drug mouse IgG levels which could be the reason for insufficient depletion of thymidine for T20 as shown in FIG. 10A.

Example 6. MNGIE Mouse Model Assay – Duration of Action

[278] To investigate the duration of action by the disclosed CPP-TP conjugates on MNGIE mice, a new group of MNGIE mice were treated with 20 mpk of the CPP-TP conjugate by intravenous injection once weekly for four weeks. Serum thymidine levels were measured on weeks one and two, 24 hours after treatment (W1-24hr and W2-24hr, respectively), on week three, 32 hours after treatment (W3-32hr), and on week four, 36 hours after treatment (W3-

36hr). 50 microliter of serum were harvested at designated time points and their levels of nucleotides were analyzed by HPLC-UV assay as described in EXAMPLE 4.

[279] The results of this assay are provided in FIG. 11. These results indicate that CPP-TP can reduce and maintain serum thymidine levels in MNGIE mice to that of healthy levels (measured for mice which do not have a mutated *TYMP* gene) for at least 32 hr.

Example 7. In Vivo Assays with PEGylated CPP-TP Conjugate

[280] Conjugation of water-soluble polymers to proteins has been reported to improve stability of the protein. Disclosed CPP-TP was PEGylated (polyethylene glycol was conjugated on CPP-TP) to investigate whether such a modification would increase cellular stability and thus increase the duration of action. PEGs having a molecular weight of 5 kDa (PEG5K), 2 kDa (PEG12), 10 kDa (PEG10K), 40 kDa with different structures (40K linear and 40K branched) were conjugated to CPP12-N-TP11, respectively, according to the following procedure.

[281] To investigate the duration of action of pegylated CPP-TP conjugates over a 48 hour period following intravenous injection, PBS control, CPP12-N-TP11, CPP12-N-TP11-AF658, CPP12-N-TP11-PEG5K or CPP12-N-TP11-PEG12 were administered to wild type CD1 mice at 8 mpk. Thymidine concentration (μM) in serum was measured at the following time points after tail-vein intravenous administration: 5 min; 8 hr; 24 hr; 36 hr; and 48 hr as described in EXAMPLE 4 (FIG. 13). These results showed that PEGylated CPP-TP conjugates maintain reduced thymidine levels for elongated amount of time compared to non PEGylated analogues. Notably, the PEGylated conjugates (e.g., cCPP12-N-TP11-PEG5K) which significantly reduced the thymidine levels for at least 48 hours following intravenous injection, whereas thymidine levels return to control levels after about 24 hours in the cases of are non-PEGylated TP.

[282] To compare the pharmacokinetic profiles of PEGylated and non-PEGylated CPP12-N-TP11, specific TP enzyme activities of serum samples collected at 5 min; 8 hr; 24 hr; 36 hr; and 48 hr were tested. These results are provided in FIG. 12. The pharmacokinetic profiles of PEGylated and non-PEGylated cCPP12-N-TP11 did not show significant difference, as all cCPP12-N-TP11 variants showed minimal TP enzyme activity in serum 24 hr post administration.

The duration of action of cCPP12-N-TP11-PEG5K was further investigated using MNGIE mice. In this case, MNGIE mice from EXAMPLE 6 were subjected to intravenous administration of cCPP12-N-TP11-PEG5K at 16 mpk. Thymidine levels in serum were

measured for the treated MNGIE mice and compared to untreated MNGIE mice. The assay was conducted according to Example 6 and additional study day on week six was added. Specifically, serum thymidine concentrations were measured on week six 48 hours after treatment (W6-48hr). The results of cCPP12-N-TP11-PEG5K are included in FIG. 14. Compared to cCPP12-N-TP11, which can reduce and maintain serum thymidine levels in MNGIE mice for approximately 36 hr (W4-36hr, FIG. 14), PEGylated cCPP-TP conjugates (e.g., cCPP12-N-TP11-5K) are able to deplete serum thymidine levels for at least 48 hours after treatment (W6-48hr, FIG. 14). To compare the efficacy of different PEGylated cCPP12-N-TP11, cCPP12-N-TP11-PEG5K, cCPP12-N-TP11-PEG10K, cCPP12-N-TP11-PEG40K linear or cCPP12-N-TP11-PEG40K branched were administered to MNGIE mice from EXAMPLE 6. MNGIE mice were treated once per week for 4 weeks. The first injection was done at 10mpk and followed by 3 more injections at 5mpk. Thymidine concentration (μM) and TP enzyme activity in serum was measured at the following time points after tail-vein intravenous administration: 3 days and 6 days for each week after one injection as described in EXAMPLE 6. Three days post injection, cCPP12-N-TP11-PEG10K, cCPP12-N-TP11-PEG40K linear and cCPP12-N-TP11-PEG40K branched treated mice showed significantly decreased Thymidine concentration at 10 mpk injections during week 1 (FIG. 23). Over four-week period, the efficacy were quantified by Thymidine or deoxyuridine level three days post injection. And the efficacy of cCPP12-N-TP11-PEG40K linear and cCPP12-N-TP11-PEG40K treatments were decreasing over multiple injections (*see* FIG. 24A and 24B). On the other hand, cCPP12-N-TP11-PEG10K significant decreased the metabolite (Thymidine) level for over four weeks (FIG. 25). In addition to the metabolite level, we also quantified the TP enzyme activity in the serum. From the samples collected during week 1, we found TP activity in the serum especially from cCPP12-N-TP-PEG40K injected mice (FIG. 26A). From the samples collected during week 4, however, the circulating TP activity from serum samples collected from cCPP12-N-TP-PEG40K injected mice dropped significantly to background level (FIG. 26B). This indicates the presence of anti-drug antibody against PEG40K modified protein over repetitive injections possibly due to its elongated half-life in the circulation.

Example 8. Pharmacokinetics of cCPP12-TP and TP Proteins

[283] The pharmacokinetics of the cCPP12-TP11 conjugates of the present disclosure or TP were investigated using mouse models. The cCPP12-TP11 conjugate (20 mpk) was intravenously injected into wild-type CD1 mice, and serum samples were collected at the following time points after administration: 0.5 hour; 2 hours; 8 hours; 24 hours; and 48 hours

(FIG. 15A). As a comparator, 20 mpk intravenous injection of free TP (not conjugated to CPP) was also injected into wild-type mice, and serum samples were collected at the same time points (FIG. 15B). Samples were analyzed and quantified using a western blot assay.

[284] The data shows that cCPP-TP11 disappear from the circulation by 12 hour post IV injection, which indicates that the half life of cCPP-TPs are significantly shorter than that of the free TP, indicating depletion from circulation and intracellular delivery of the cargo TP protein in a cCPP-dependent manner.

[285] To compare the pharmacokinetic and pharmacodynamic profiles of the cCPP12-N-TP11 and cCPP12-N-TP16, wild type mice were injected with 5 mpk cCPP12-N-TP11 or cCPP12-N-TP16, or PBS control, and serum samples were collected 5 min, 8 hr, 24 hr, 36 hr, or 48 hr post injection. Specific TP enzyme activities of serum samples were measured by enzyme activity assay described above and thymidine level of serum samples were also tested. These results are provided in FIG.27A and 27B.

[286] The data showed that the pharmacokinetic profiles of cCP12-N-TP11 and cCPP12-N-TP16 did not show significant difference. And both cCPP12-N-TP11 and cCPP12-N-TP16 variants showed comparable efficacy in reducing the thymidine levels in vivo.

[287] To compare the pharmacokinetic and pharmacodynamic profiles of cCPP12-N-TP11-PEG10K and cCPP12-N-TP16-PEG10K, wild type mice were injected with 5 mpk cCPP12-N-TP11-PEG10K or cCPP12-N-TP16-PEG10K. Serum samples were collected at 5min, 8 hr, 24 hr, 48 hr, 72 hr or 96 hr post injection. Specific TP enzyme activities of serum samples were measured by enzyme activity assay described above, and thymidine level of serum samples were also tested as described above. The results were provided in FIG. 28A and 28B.

[288] The data showed that similar to the non-PEGylated protein, PEGylated cCPP-TP11 and cCPP-TP16 did not show significant difference for the pharmacokinetic profiles. And both cCPP12-N-TP11-PEG10K and cCPP12-N-TP16-PEG10K showed effects for the depletion of thymidine level in vivo.

[289] To compare the pharmacokinetic and pharmacodynamic profiles of cCPP12-SS-TP11-PEG10K and cCPP12-SS-TP16-PEG10K, wild type mice were injected with 5mpk cCPP12-SS-TP11-PEG10K or cCPP12-SS-TP16-PEG10K, and serum samples were collected 5min, 8 hr, 24 hr, 48 hr, 72 hr or 96 hr post injection. Specific TP enzyme activities of serum samples were measured by enzyme activity assay described above, and thymidine level of serum samples were also tested as described above. The results are provided in FIG. 29A and 29B.

[290] The data showed that cCPP12-SS-TP11-PEG10K showed similar pharmacokinetic profiles compared to cCPP12-SS-TP16-PEK10K. Both TP derivatives showed effects for the depletion of thymidine level in vivo.

[291] To compare the pharmacokinetic and pharmacodynamic profiles of Fc₋P16 and cCPP12-N-Fc-TP16, wild type mice were injected with Fc-TP16 and cCPP12-N-Fc-TP16 at 5 mpk, and serum samples were collected at 5 min, 24 hr, 48 hr, 72 hr, 120 hr or 156 hr post injection. Specific TP enzyme activities of serum samples were measured by enzyme activity assay described above, and thymidine level of serum samples were also tested as described above. The results are provided in FIG. 32 and 33.

Example 9. Biodistribution

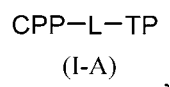
[292] The biodistribution of the CPP-TP conjugates in the wild type CD1 mice 4 h, 8 h, or 24 h after intravenous administration was investigated using a fluorescently labeled CPP-TP. The CPP-TP was fluorescently labeled with AlexaFluor568 (AF568) and yielded the CPP-TP-AF568 used in this study. Briefly, CPP12-N-TP (2 mg/mL, 40 μ M) was mixed with the AlexaFluoro568-NHS ester (NHS, N-hydroxyl succinimide, ThermoFisher) at 1:8 molar ratio in phosphate buffer saline (pH 7.4). To separate extra fluorophore, PD-10 columns were applied according manufacturer's protocol and CPP-TP-AF568 were obtained. Combined fractions with desired product were sterile filtered, concentrated, and stored at 2 mg/mL at -20 °C.

[293] To study the tissue distribution of CPP-TP, mice were injected intravenously at 5 mpk with CPP-TP-AF568. Then, mice were anesthetized, bled, euthanized, and dissected 4 h, 8 h, or 24 h after injection. Heart, kidney, liver, lung, large intestine, small intestine, and spleen were harvested; each piece was weighed, and several organs were halved for cryosection and tissue homogenization. For fluorescence quantification, organs were homogenized with a Tissue Lyser II system in pre-chilled tubes, stainless steel beads, and RIPA buffer supplemented with 1x protease inhibitor. Supernatant after centrifuge were obtained and transferred for fluorescence quantification. Tissues harvested from uninjected mice were used as blank and were also spiked with known concentration of CPP-TP-AF568 to generate a standard curve. The intensity of the fluorescence of samples and standard probe was detected at fluorescence plate reader. The concentrations of the homogenates were extrapolated from the calibration curves made for each organ. The tissue concentration were back-calculated with the dilution factor of tissue homogenization and homogenate concentrations. The results of the

biodistribution studies are presented in FIG. 16A. These results indicate that CPP-TP can internalize into wide range of cells and is predominately localized in the liver, kidney and spleen (FIG. 16A). Confocal images of the distribution of CPP-TP in the liver, lung, and intestine were obtained using confocal imaging (FIG. 16B).

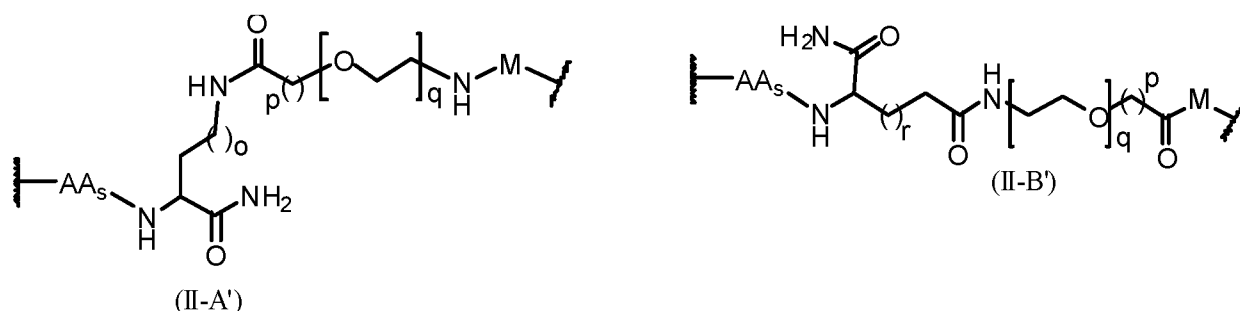
Claims

1. A compound comprising:
 - a) at least one a cell-penetrating peptide (CPP); and
 - b) a thymidine phosphorylase, or an active fragment or analog thereof (TP),
 wherein the CPP is coupled, directly or indirectly, to TP.
2. The compound of claim 1, wherein the CPP is conjugated, directly or indirectly, to the TP.
3. The compound of claim 2, wherein the CPP is conjugated to the N-terminus or the C-terminus of the TP.
4. The compound of claim 2, wherein the CPP is conjugated to a side chain of an amino acid in the TP.
5. The compound of any of claims 2-4, further comprising a linker (L), which conjugates the CPP to the TP.
6. The compound of claim 5, wherein the L is covalently bound to the N-terminus of TP.
7. The compound claim 5, wherein the L is covalently bound to the C-terminus of TP.
8. The compound of claims 5, wherein the L is covalently bound to a side chain of an amino acid on TP.
9. The compound of claim 5, having a structure according to Formula I-A:



wherein L is a covalently bound to the side chain of an amino acid on the CPP and to the N-terminus of the TP.

10. The compound of any of claims 1-9, wherein the CPP is a cyclic cell-penetrating peptide (cCPP).
11. The compound of any of claims 5-10, wherein L is one or more D or L amino acids, each of which is optionally substituted; alkylene, alkenylene, alkynylene, carbocyclyl, or heterocyclyl, each of which is optionally substituted; or $-(R^1-X-R^2)_z-$, wherein each of R^1 and R^2 , at each instance, are independently selected from alkylene, alkenylene, alkynylene, carbocyclyl, and heterocyclyl, each X is independently NR^3 , $-NR^3C(O)-$, S, and O, wherein R^3 is independently selected from H, alkyl, alkenyl, alkynyl, carbocyclyl, and heterocyclyl, each of which is optionally substituted, and z is an integer from 1 to 20; or combinations thereof.
12. The compound of any of claims 5-11, wherein L has a structure according to Formula II-A' or II-B':



and wherein

M is absent or a group bound to TP;

AA_s is a side chain or terminus of an amino acid on the CPP;

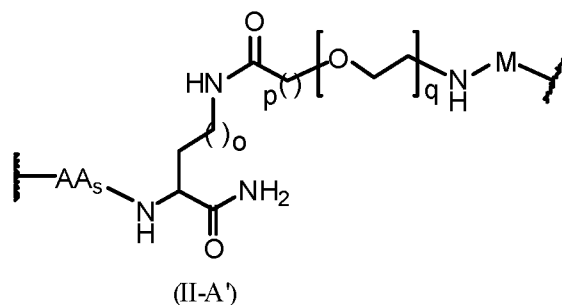
o is an integer from 0 to 10;

p is an integer from 0 to 10;

q is an integer from 1 to 50; and

r is 0 or 1.

13. The compound of claim 12, wherein L has a structure according to Formula II-A':



, and

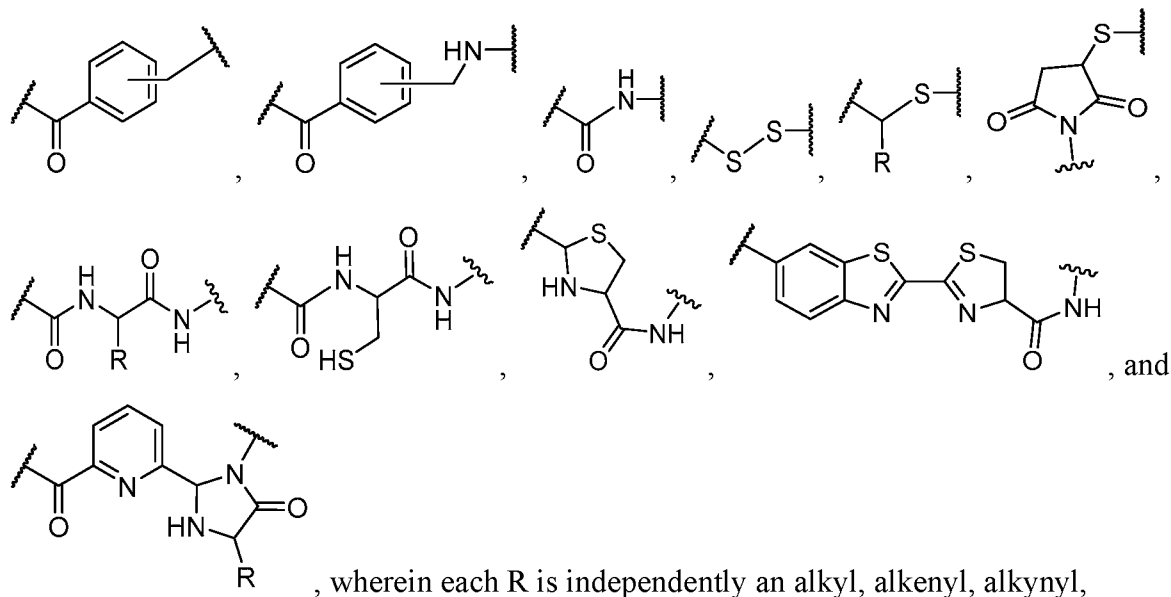
wherein o is 3;

p is 2; and

q is an integer from 10 to 15.

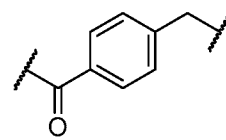
14. The compound of claim 12 or 13, wherein M is present and comprises an alkylene, alkenylene, alkynylene, carbocyclyl, or heterocyclyl, each of which is optionally substituted.

15. The compound of any of claims 12-14, wherein M is present and selected from the group consisting of:

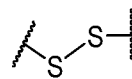


, wherein each R is independently an alkyl, alkenyl, alkynyl, carbocyclyl, or heterocyclyl.

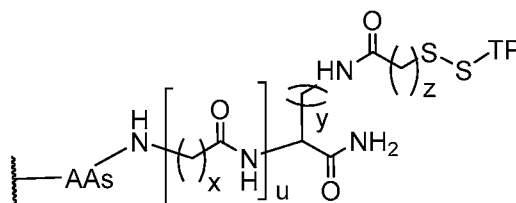
16. The compound of any of claims 12-15, wherein M is



17. The compound of any of claims 1-11 or 15, wherein M is



18. The compound of claim 17, wherein L-M has a



II-C'

wherein:

AAs is a side chain or terminus of an amino acid on the CPP;

z is an integer from 0 to 10;

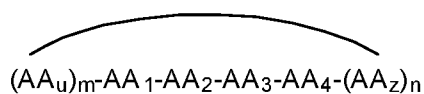
y is an integer from 0 to 10;

x is an integer from 0 to 10; and

u is an integer from 1 to 50.

19. The compound of claim 18, wherein Z is 2, y is 4, x is 2, and u is 2.

20. The compound of any of claims 10-19, wherein the cCPP has a sequence comprising Formula III:



III

wherein:

each of AA₁, AA₂, AA₃, and AA₄, are independently selected from a D or L amino acid,

each of AA_u and AA_z , at each instance and when present, are independently selected from a D or L amino acid, and

m and n are independently selected from a number from 0 to 6; and

wherein:

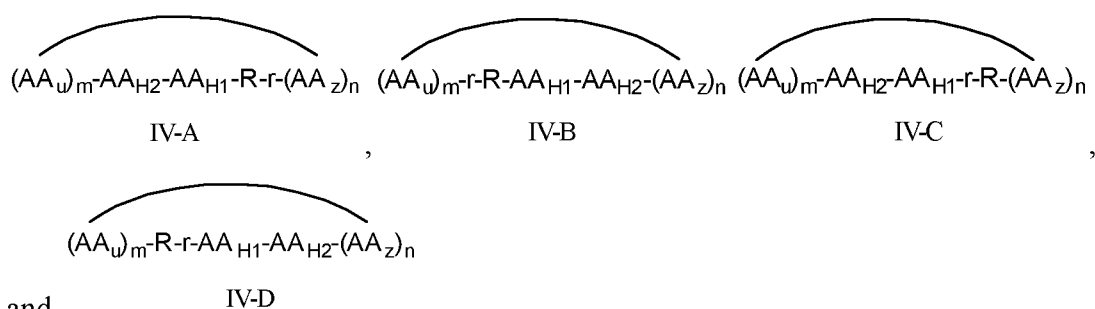
at least two of AA_u , at each instance and when present, AA_1 , AA_2 , AA_3 , AA_4 , and

AA_z , at each instance and when present, are independently arginine, and

at least two of AA_u , at each instance and when present, AA_1 , AA_2 , AA_3 , AA_4 , and

AA_z , at each instance and when present, are independently a hydrophobic amino acid.

21. The compound of 20, wherein the cCPP has a sequence comprising any of Formula IV-A-D:



wherein:

each of AA_{H1} and AA_{H2} are independently a D or L hydrophobic amino acid;

at each instance and when present, each of AA_u and AA_z are independently a D or L amino acid; and

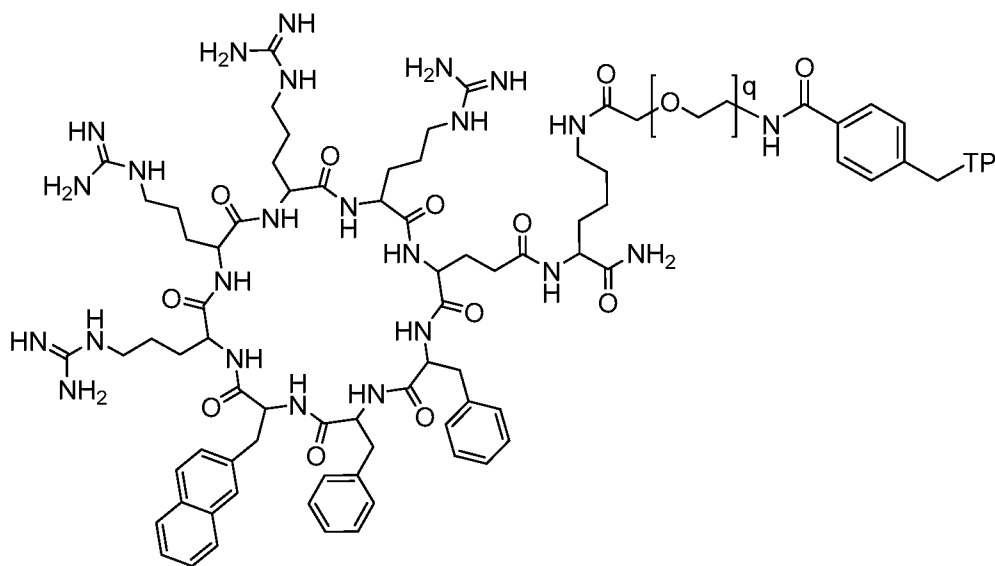
m and n are independently selected from a number from 0 to 6.

22. The compound of any of claims 1-21, wherein M is present and covalently bound to the N-terminus of TP or the C-terminus of TP.

23. The compound of any of claims 1-22, wherein M is present and is covalently bound to the N-terminus TP.

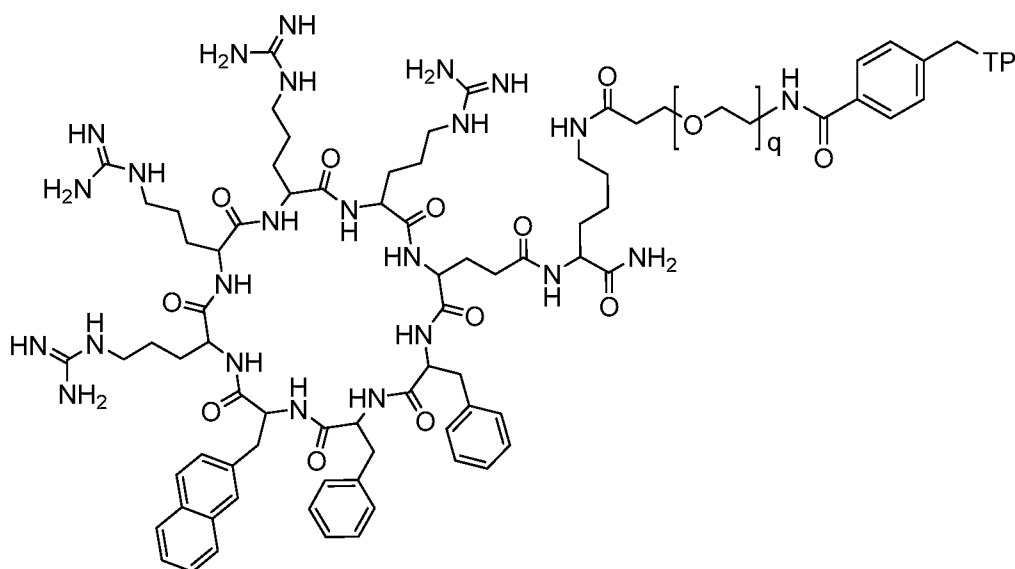
24. The compound of any of claims 1-23, wherein M is present and is covalently bound a side chain of an amino acid in TP.

25. The compound of any of claims 12-24, wherein u is 0.
26. The compound of any of claims 12-25, wherein p is 1.
27. The compound of any of claims 12-26, wherein q is 12.
28. The compound of any of claims 1-27, wherein the compound has a structure according to Formula V-A1, V-A2, or V-A3:

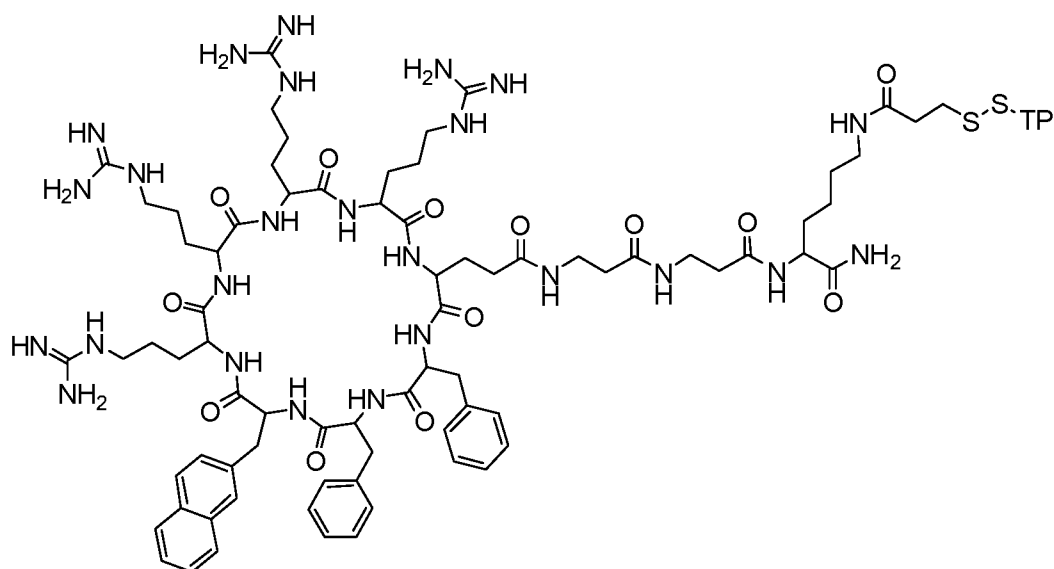


(V-A1)

or

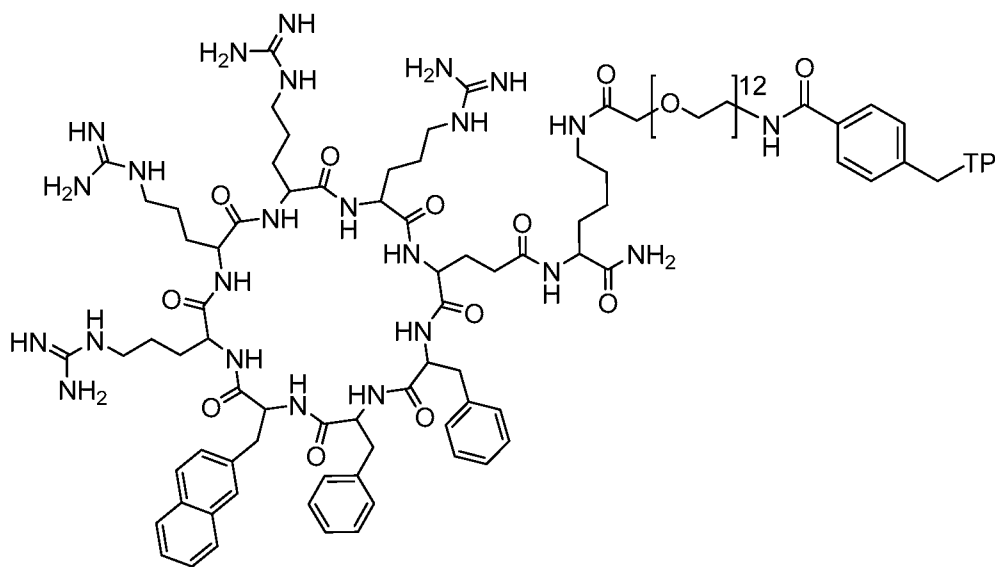


(V-A2)



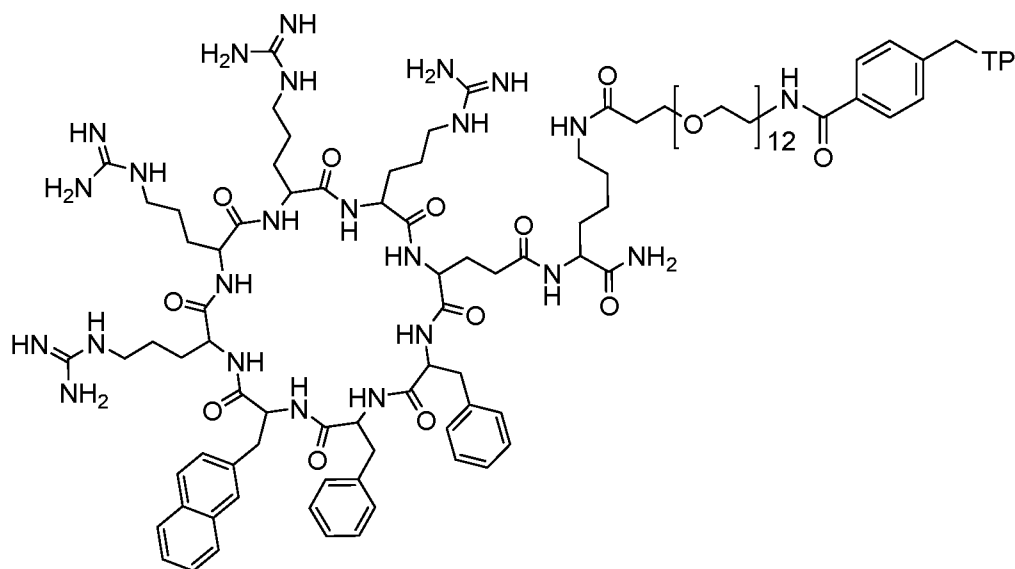
(V-A3)

29. A compound according to any of claims 1-28, having a structure according to Formula V-B1 or V-B2:



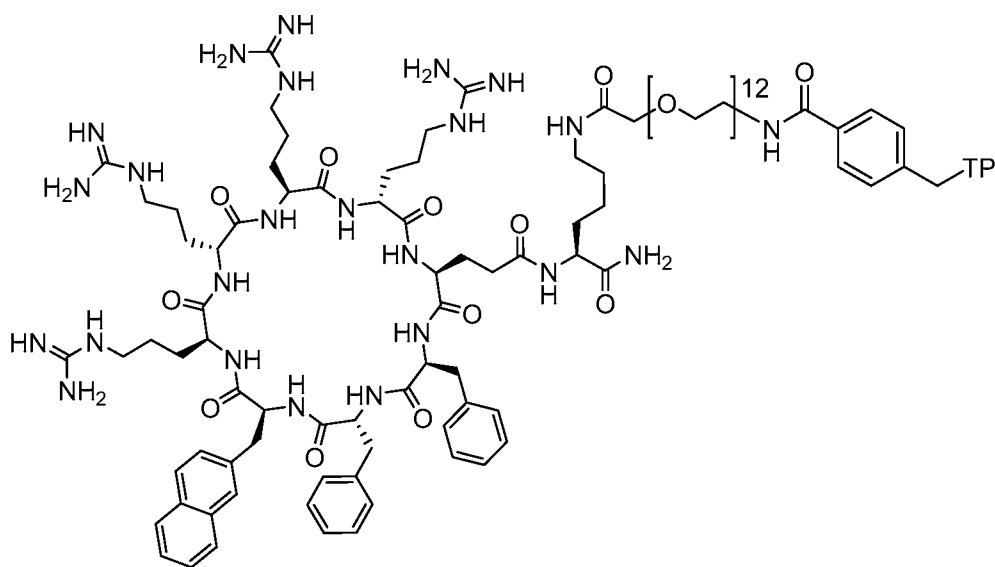
(V-B1)

or



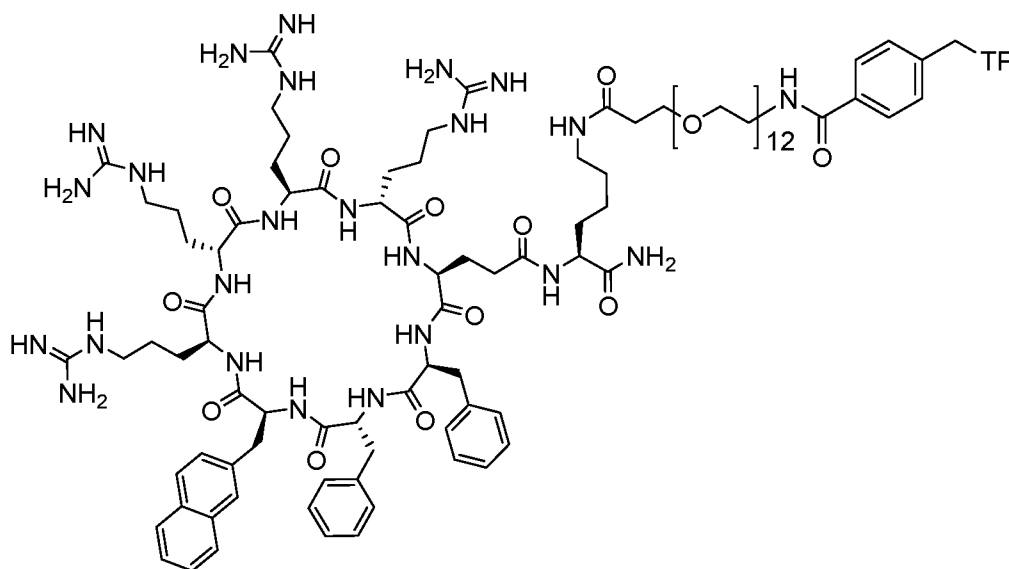
(V-B2)

30. A compound according to any of claims 1-29, having a structure according to Formula V-B3 or V-B4:



(V-B3)

or



(V-B4)

31. The compound of any of claims 1-30, wherein the TP has an amino acid sequence that is at least 95% identical to SEQ ID NO. 1.
32. The compound of any of claims 1-31, wherein TP has an amino acid sequence that is at least 95% identical to SEQ ID NO. 2.
33. The compound of any of claims 1-32, wherein TP has an amino acid sequence that is at least 95% identical to SEQ ID NO. 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13.
34. The compound of any of claims 1-33, wherein TP has an amino acid sequence that is at least 95% identical to SEQ ID NO. 3.
35. The compound of any of claims 1-34, further comprising a water-soluble polymer conjugated the TP.
36. The compound of claim 35, wherein the water-soluble polymer comprises a polyethylene glycol (PEG) residue.
37. The compound of claim 36, wherein the PEG residue has a molecular weight ranging from about 1 kDa to about 100 kDa.

38. The compound of claim 36, wherein the PEG residue has a molecular weight ranging from about 1 kDa to about 20 kDa.
39. The compound of claim 38, wherein the PEG residue has a molecular weight of about 10 kDa.
40. A method of treating Mitochondrial Neurogastrointestinal Encephalopathy in a patient in need thereof, comprising administering a compound of any of claims 1-39.
41. A method of reducing extracellular and intracellular levels of thymidine in a patient in need thereof, comprising administering a compound of any of claims 1-340.
42. The method of claim 41, for treating Mitochondrial Neurogastrointestinal Encephalopathy.
43. A cell comprising a compound of any of claims 1-41.
44. A pharmaceutical composition comprising the compound of any of claims 1-41.

FIG. 1A

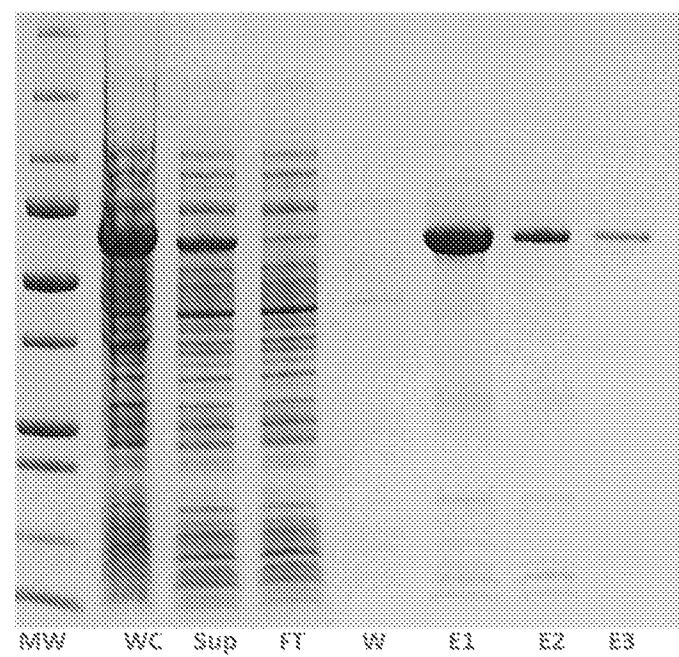


FIG. 1B

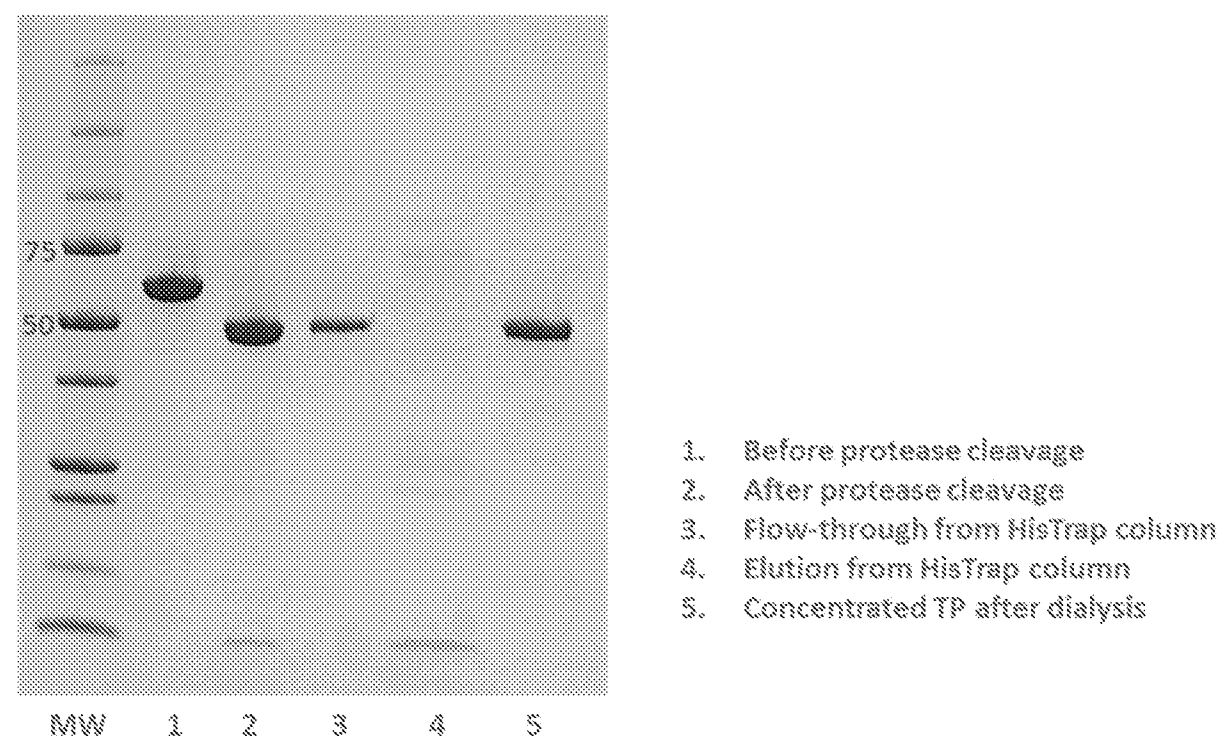


FIG. 2A

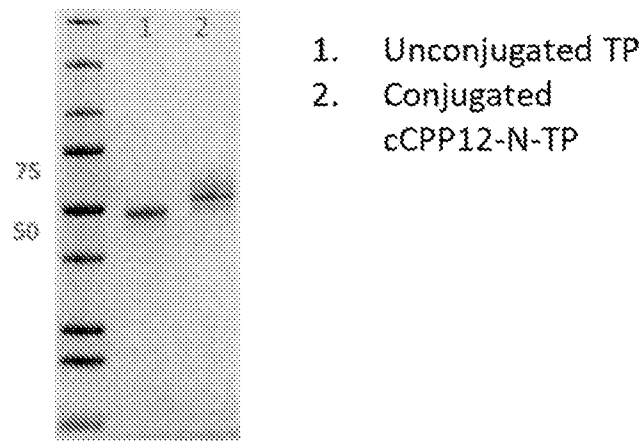


FIG. 2B

Enzyme stability

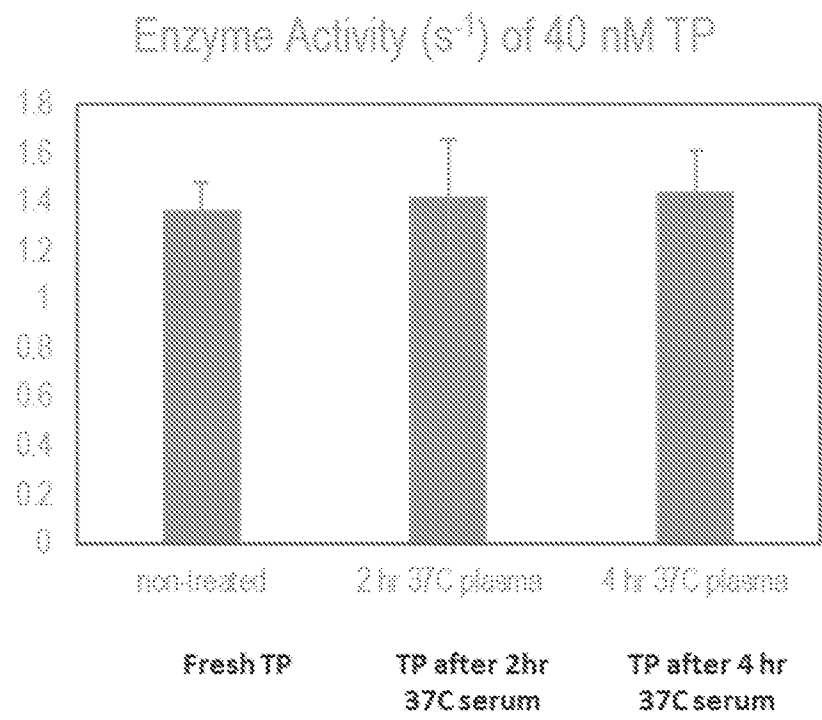


FIG. 3

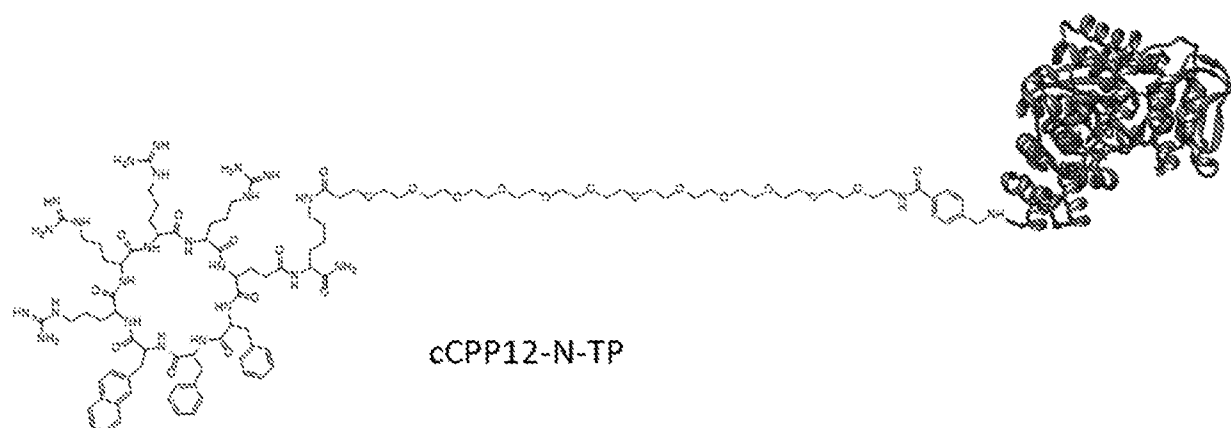
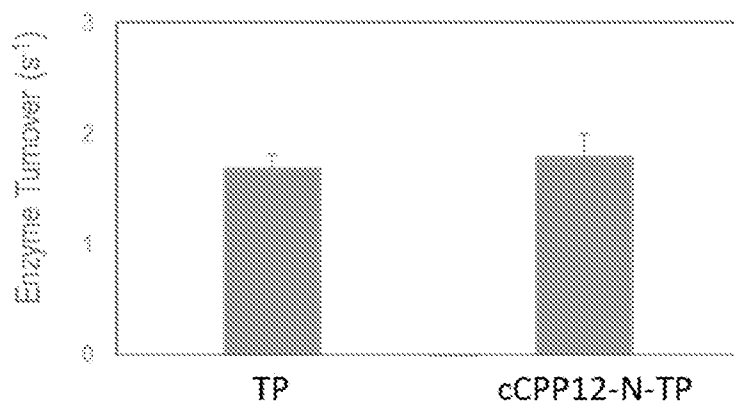


FIG. 4A

Comparison of Enzyme Activity of 40 nM of
TP and CPP-TP



4/27

FIG. 4B

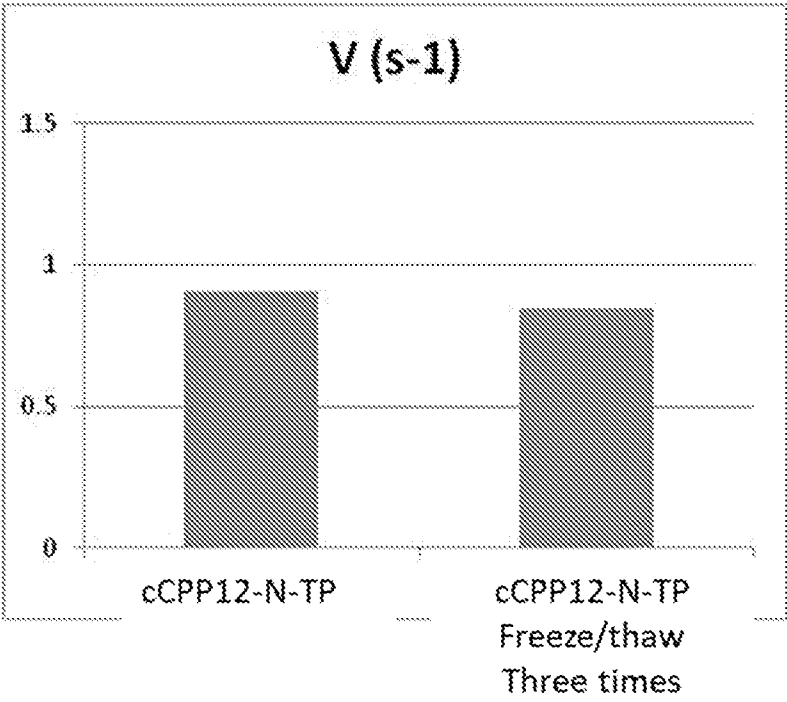
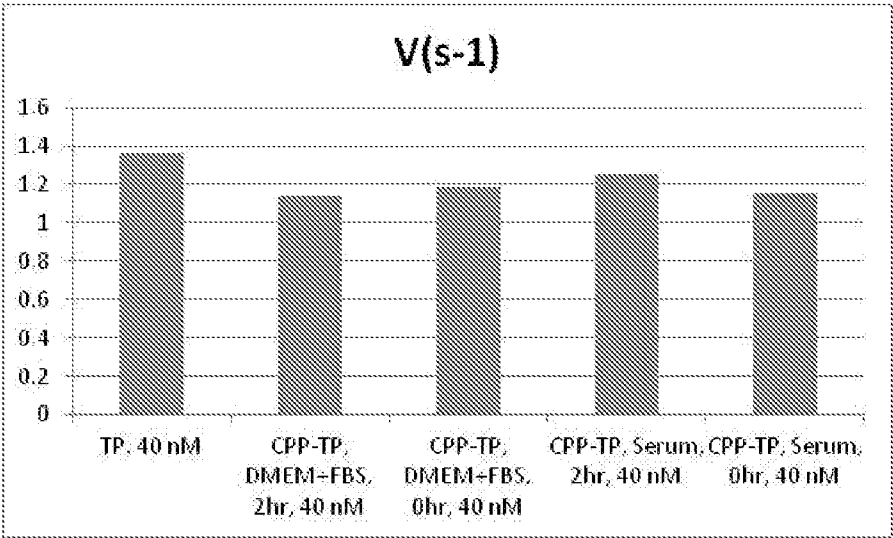
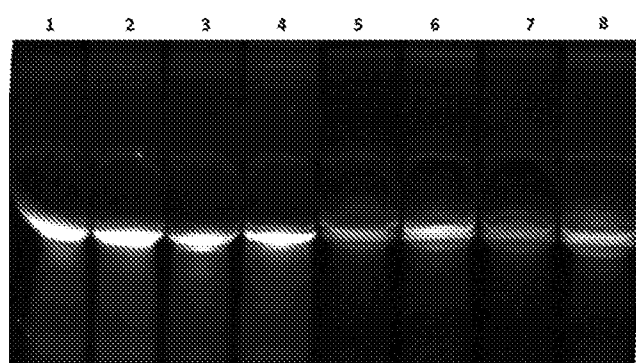


FIG. 4C



5/27

FIG. 5



Lanes:

1. TP-Alexa568 0 hr
2. TP-Alexa568 2 hr
3. TP-Alexa568 12 hr
4. TP-Alexa568 24 hr
5. CP-N-TP-Alexa568 0 hr
6. CP-N-TP-Alexa568 2 hr
7. CP-N-TP-Alexa568 12 hr
8. CP-N-TP-Alexa568 24 hr

FIG. 6A

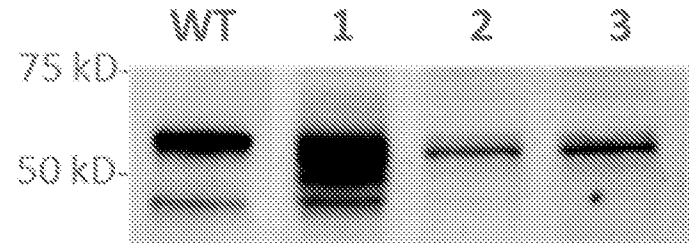
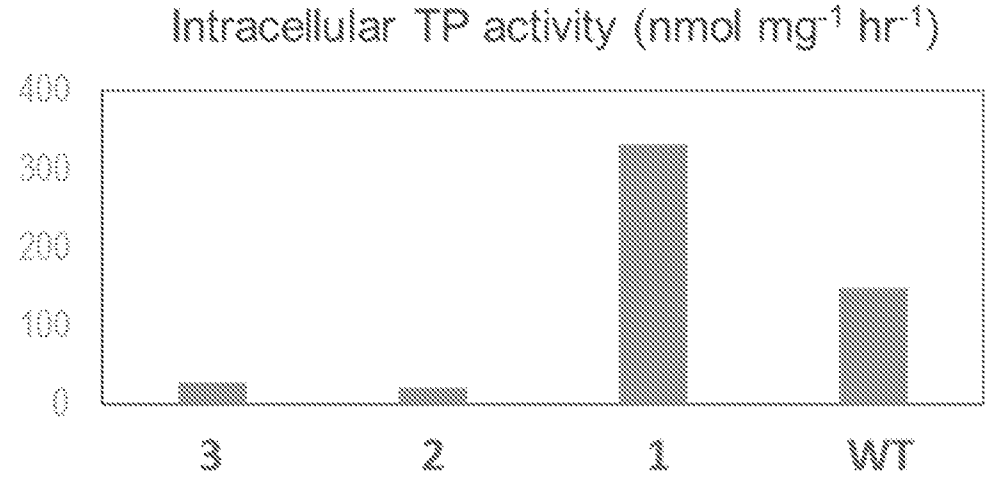


FIG. 6B



7/27

FIG. 7

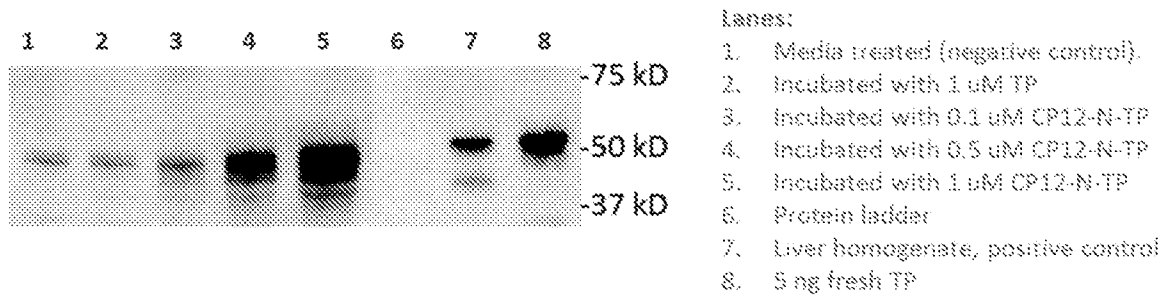


FIG. 8

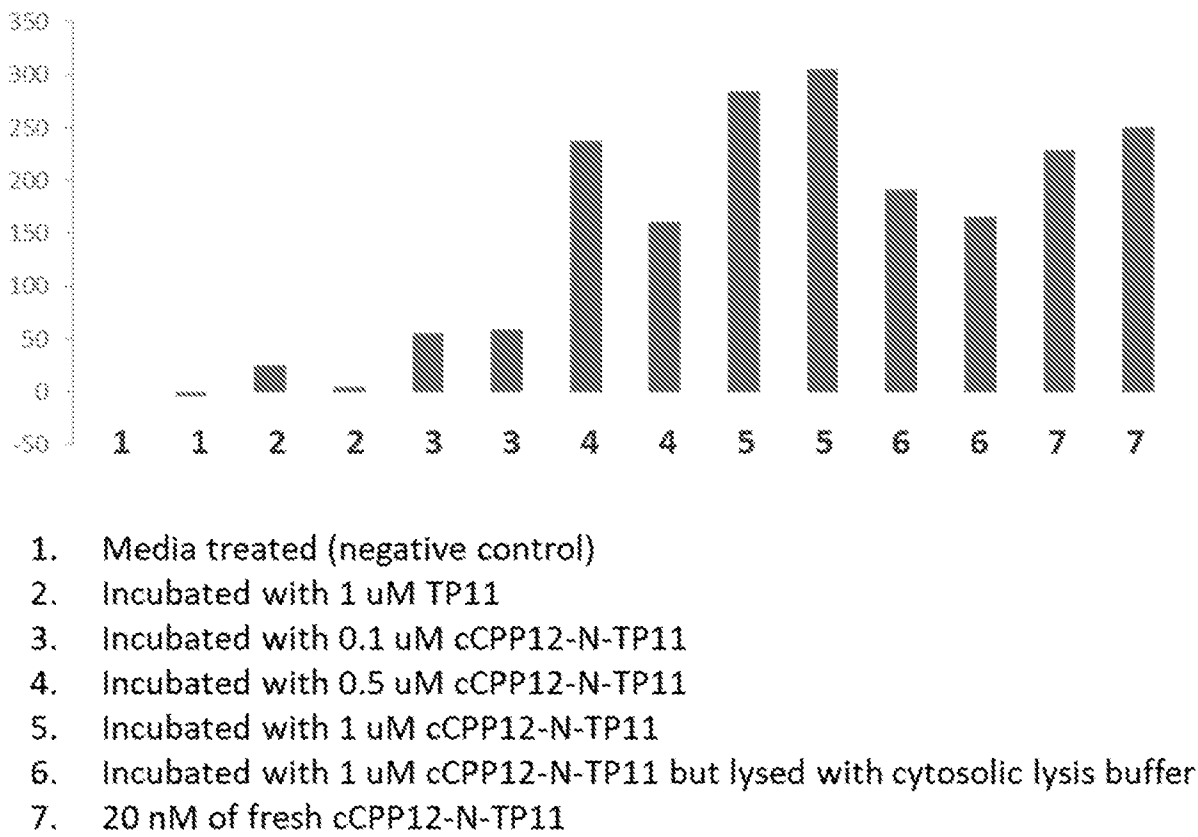
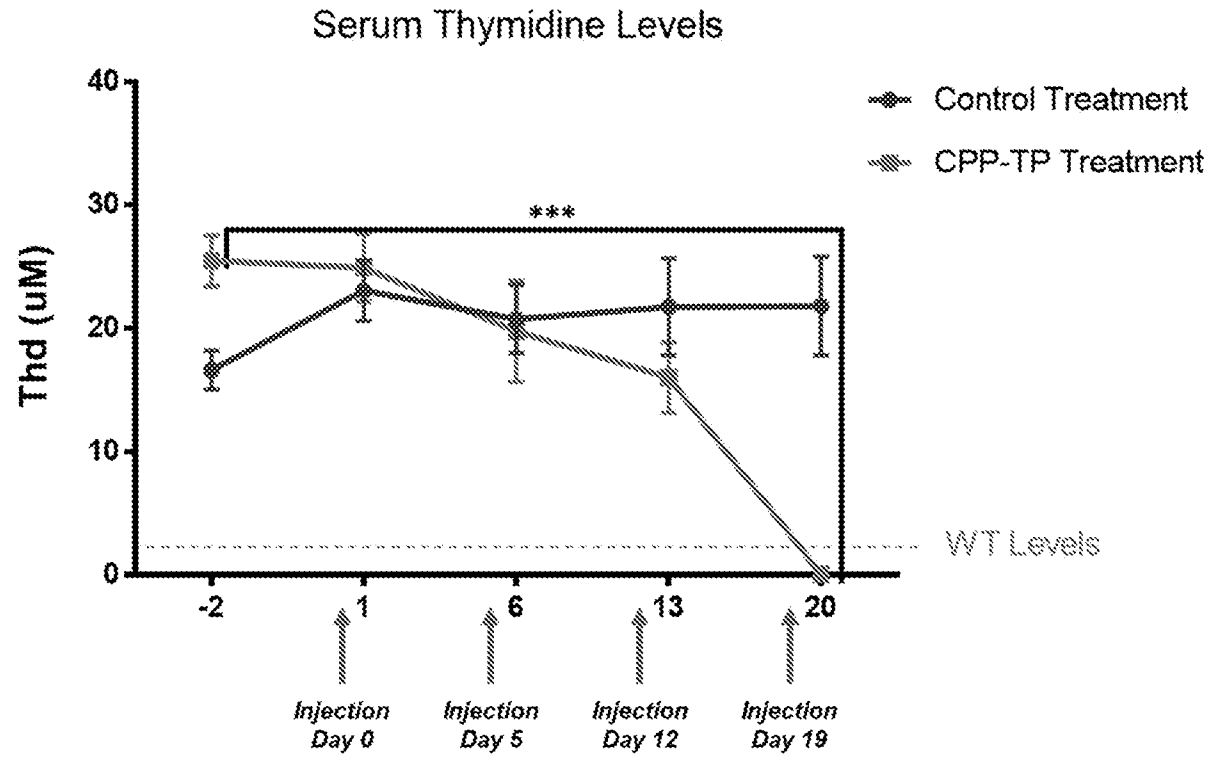


FIG. 9



Treatment 7.5 mpk on day 0 and 5, and 20 mpk on day 12 and 19
One-way ANOVA with Geisser-Greenhouse correction. *** $p < 0.0001$

9/27

FIG. 10A

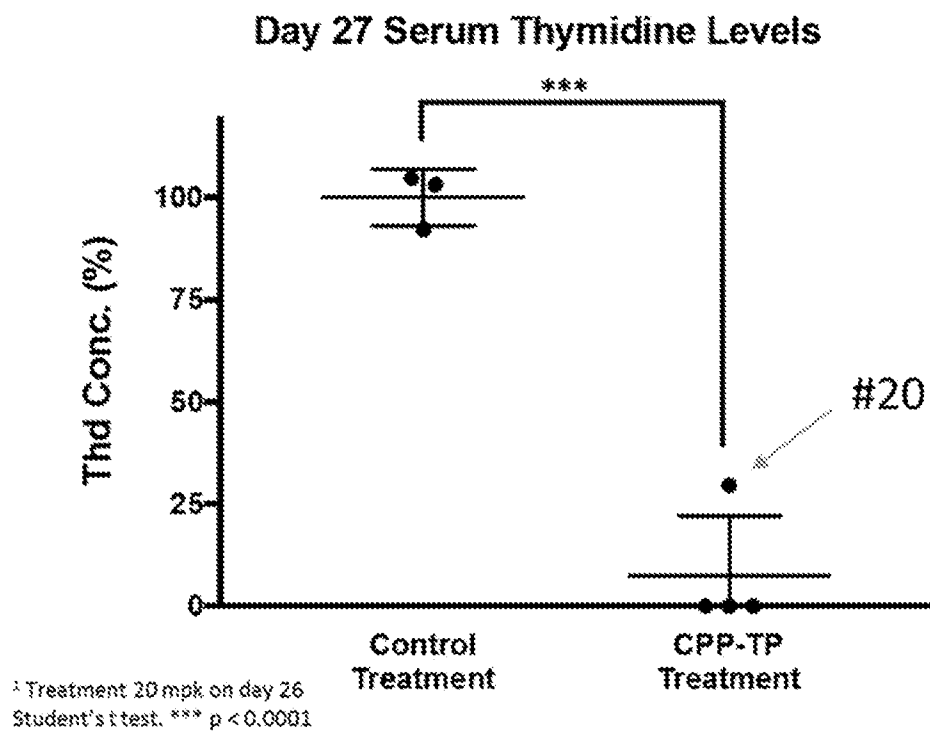
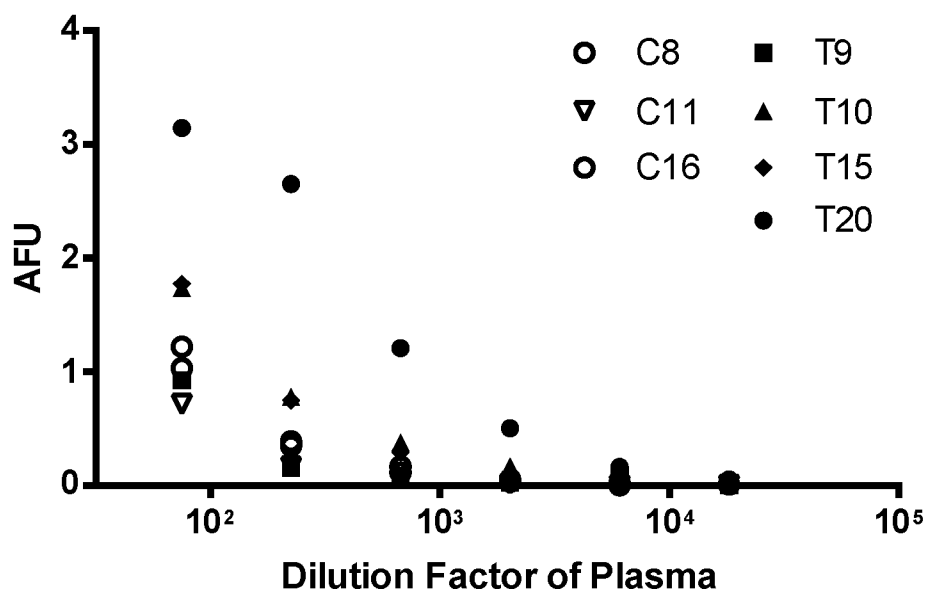


FIG. 10B

Detection of Anti-CPP-TP Antibody by ELISA

10/27

FIG. 11

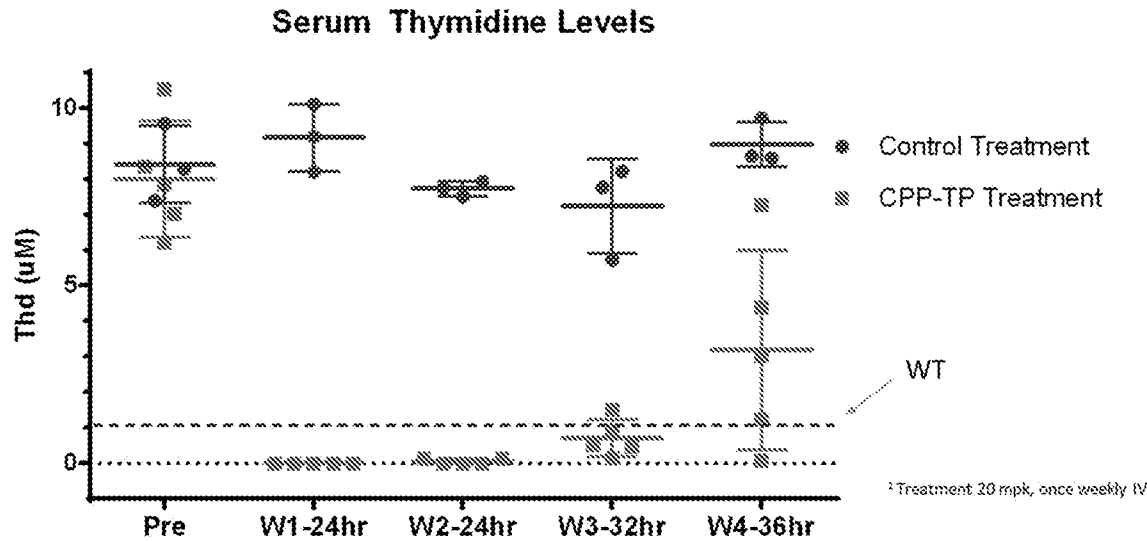
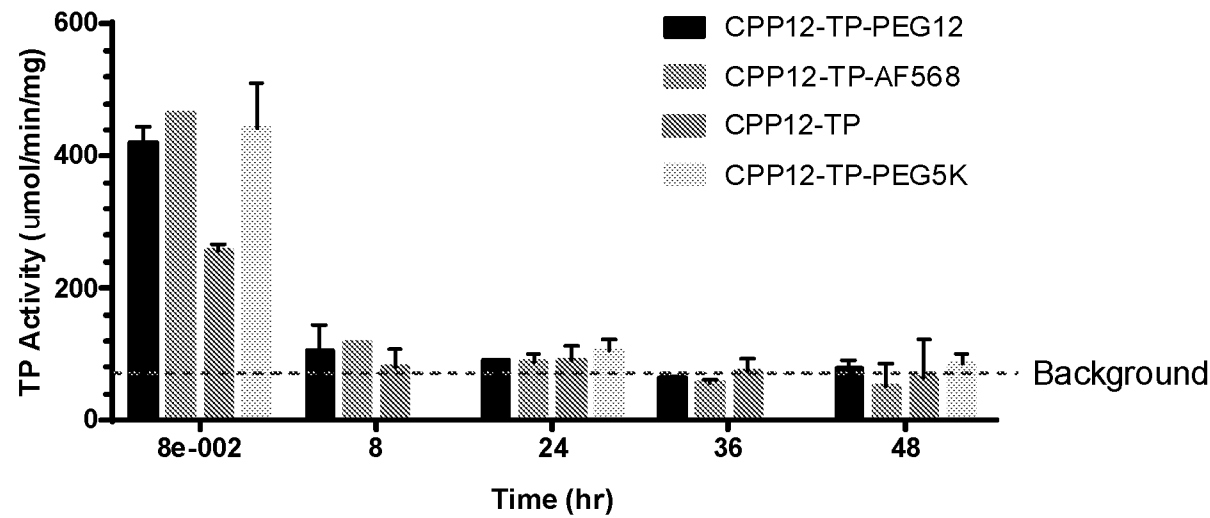
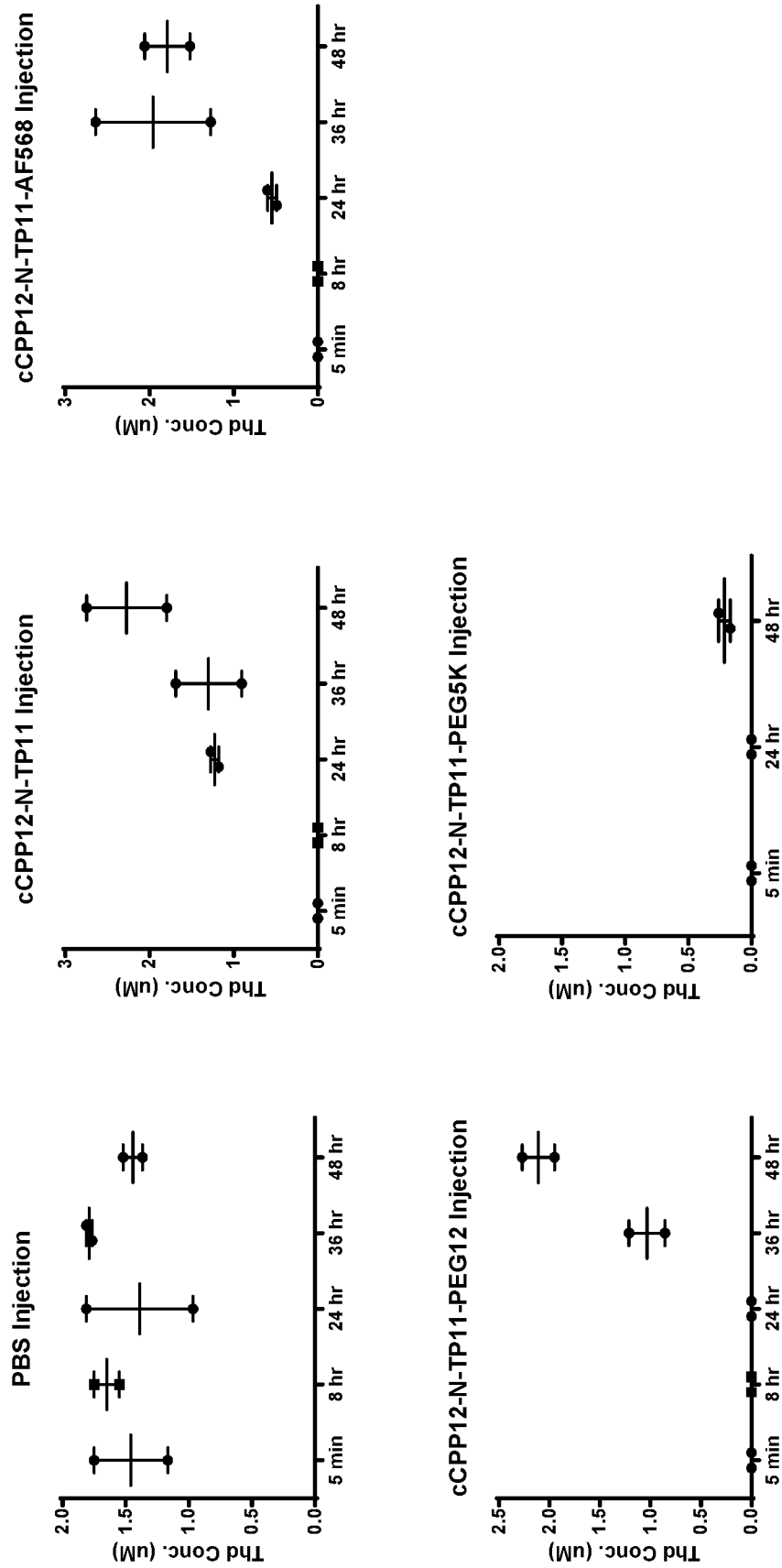


FIG. 12



11/27

FIG. 13



12/27

FIG. 14

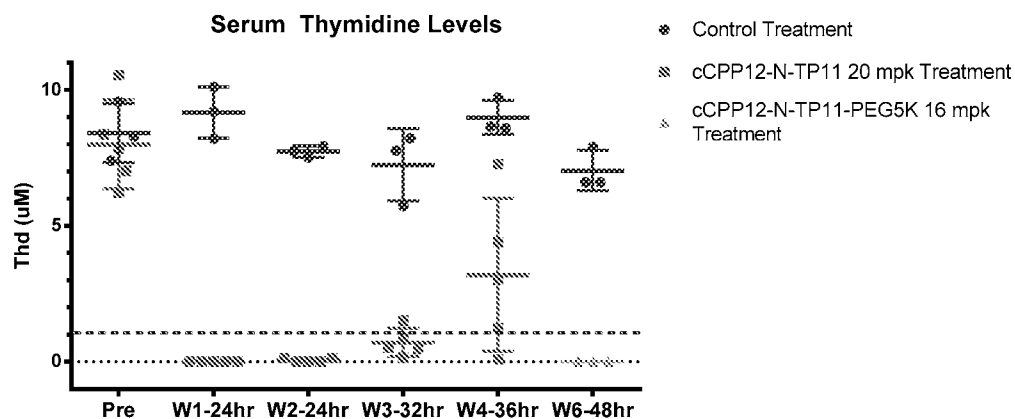


FIG. 15A

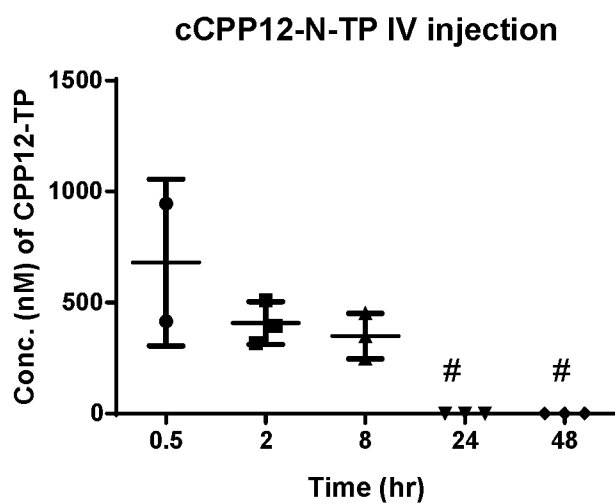


FIG. 15B

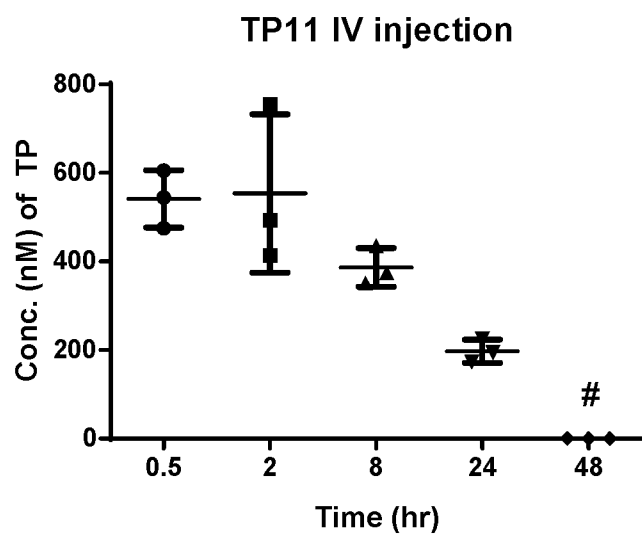


FIG. 16A

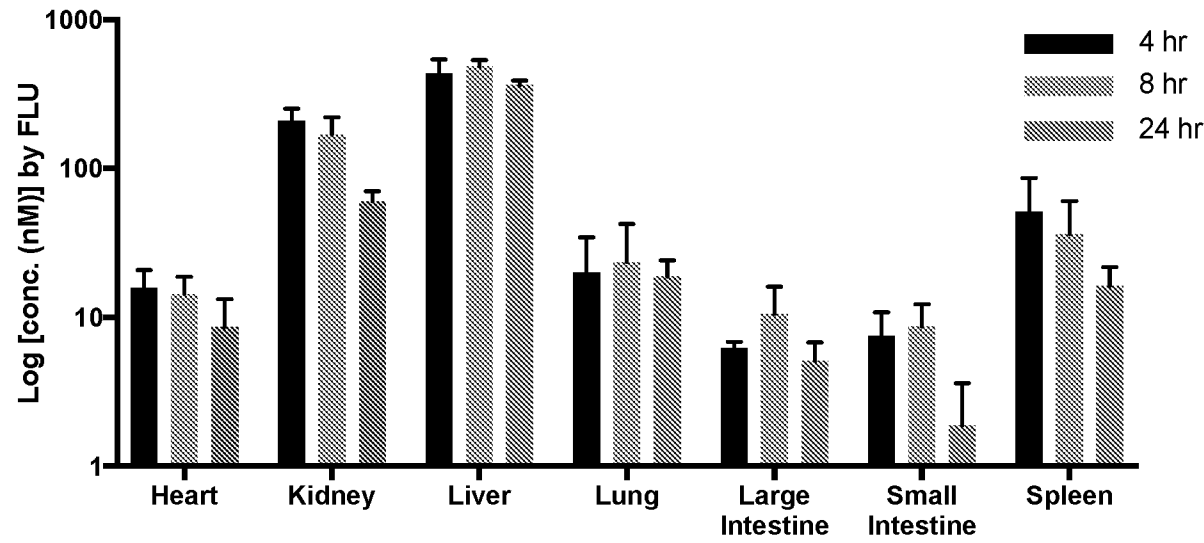
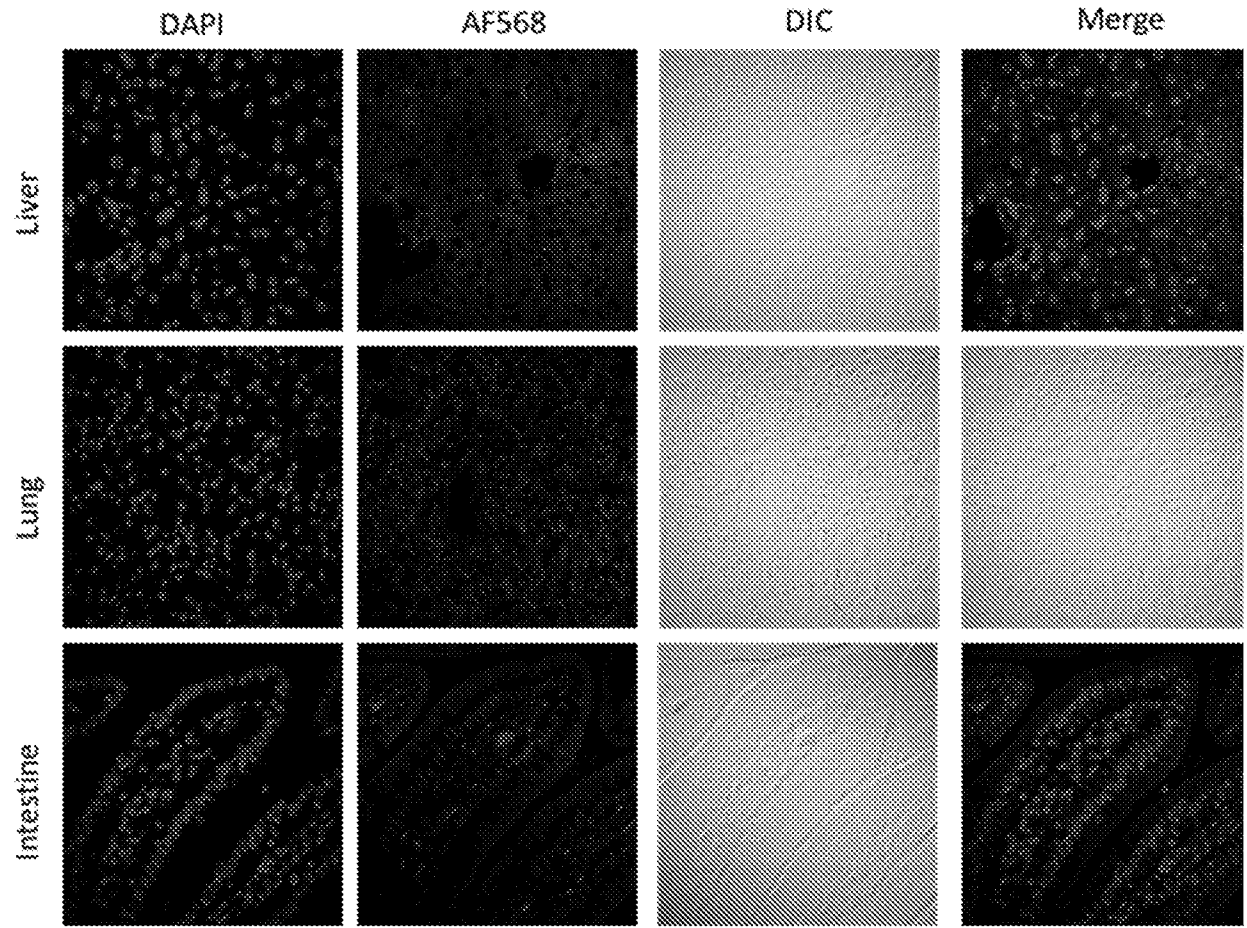


FIG. 16B



14/27

FIG. 17A.

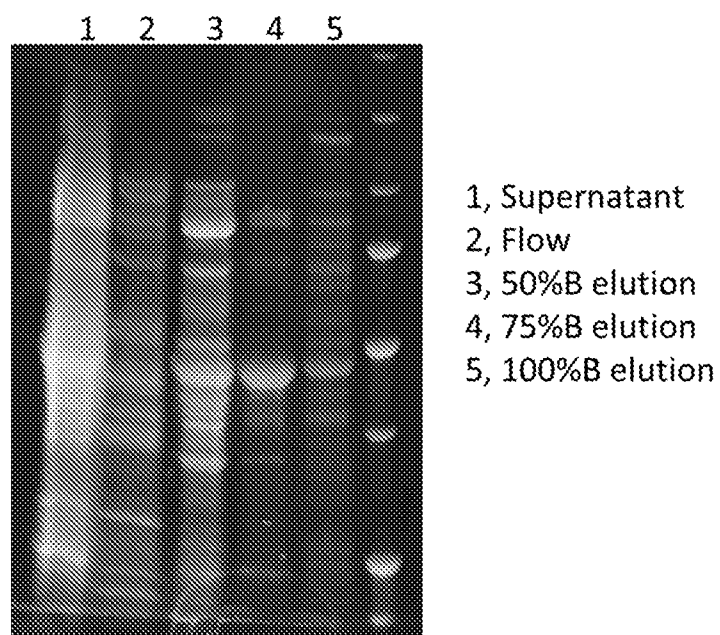


FIG. 17B

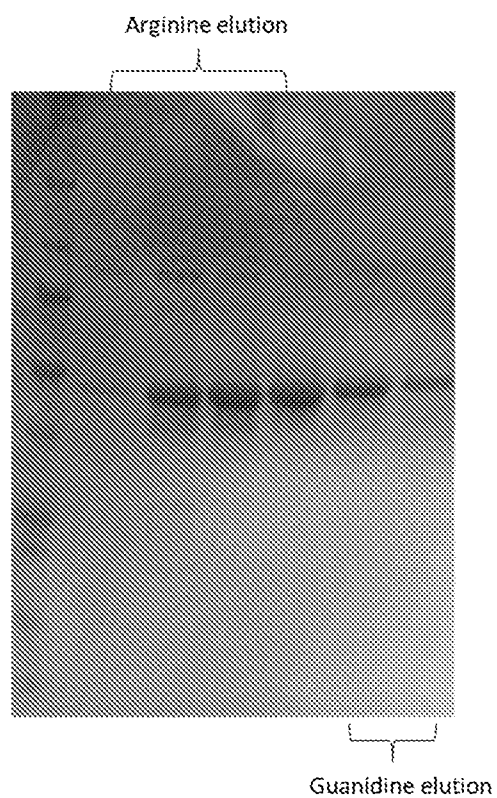


FIG. 18

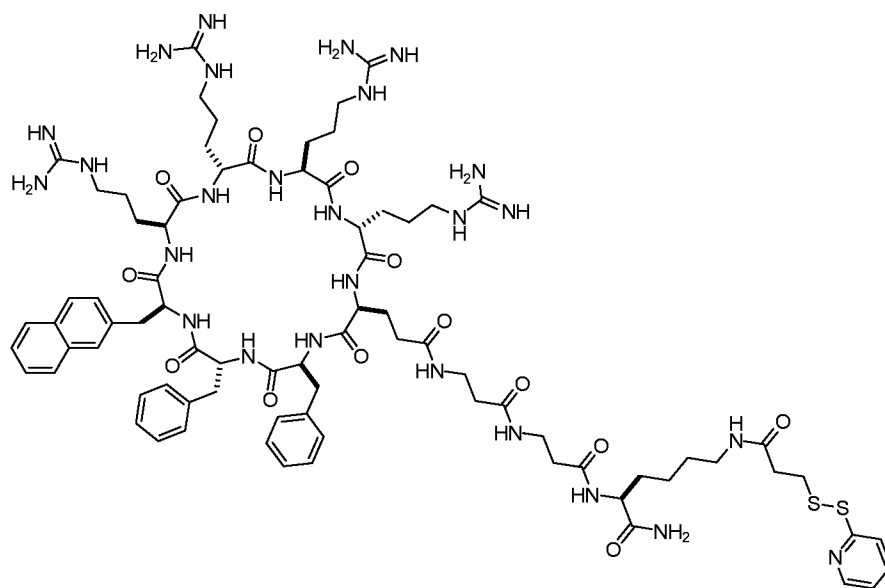
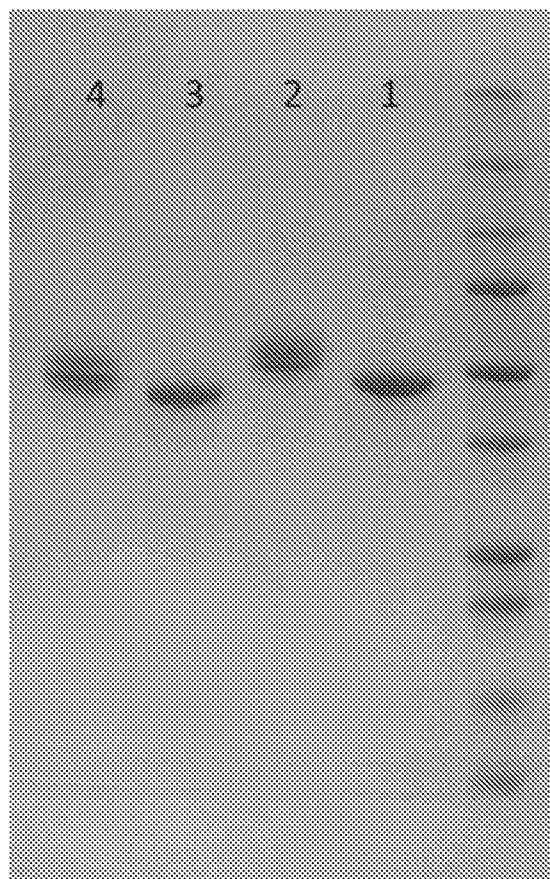


FIG. 19



1. TP11
2. cCPP12-N-TP11
3. TP16
4. cCPP12-N-TP16

16/27

FIG. 20

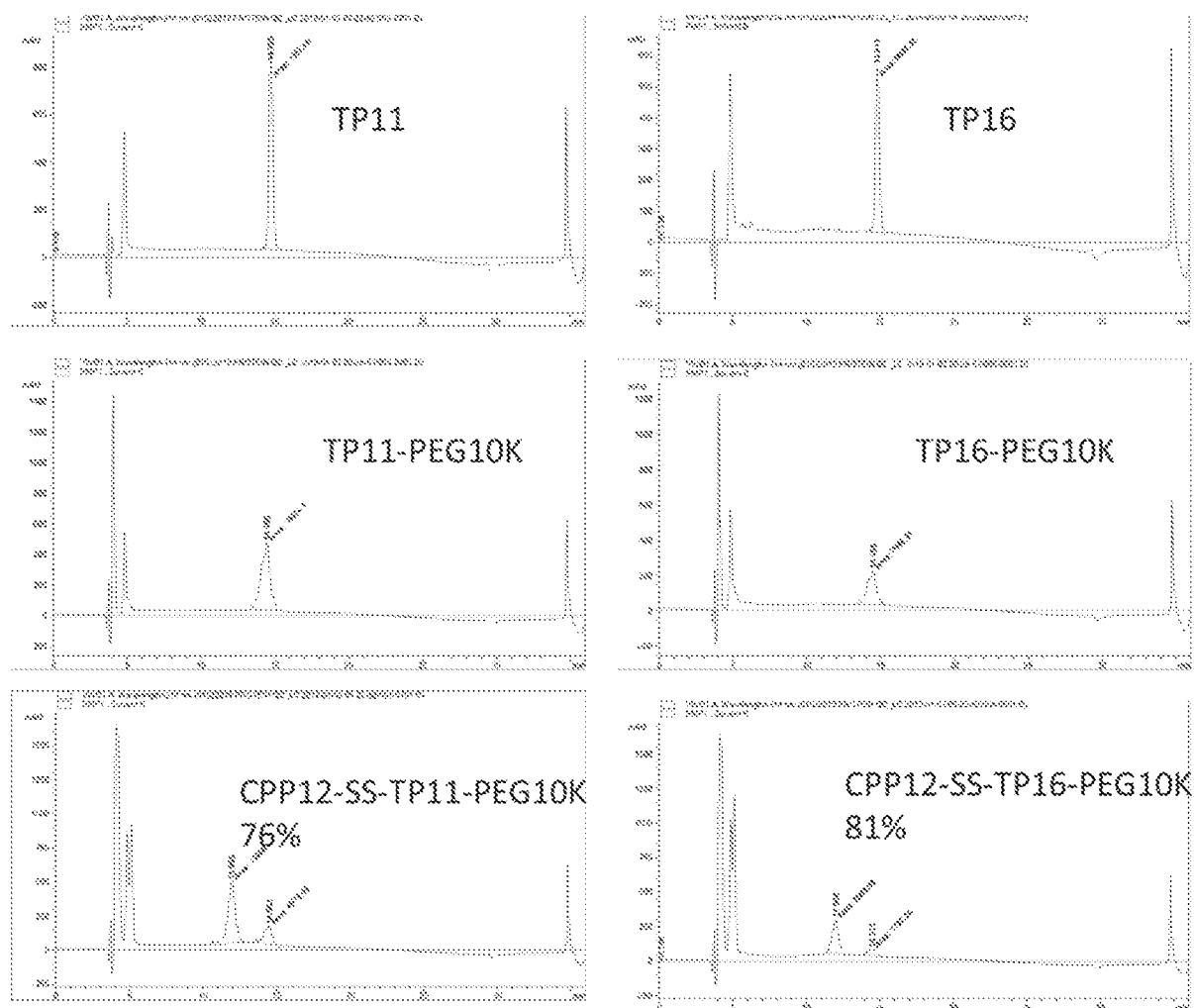
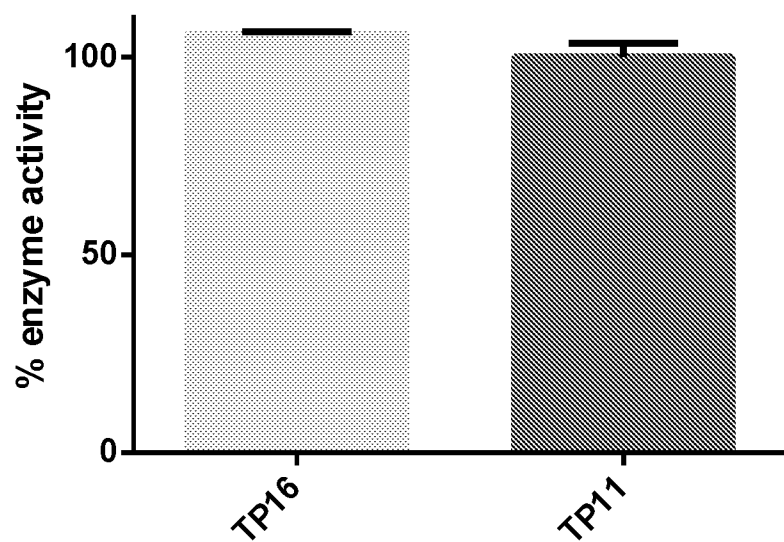


FIG. 21



17/27

FIG. 22

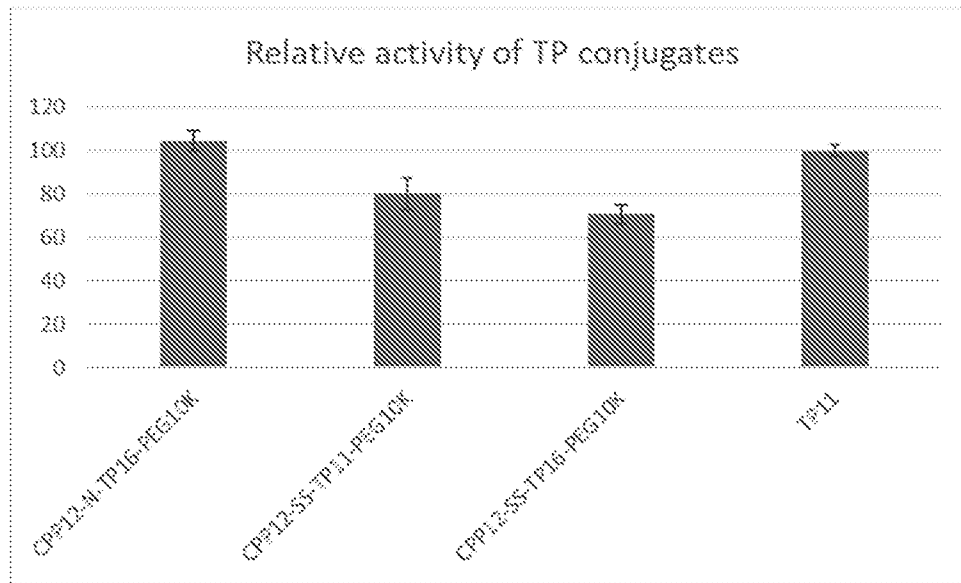


FIG. 23

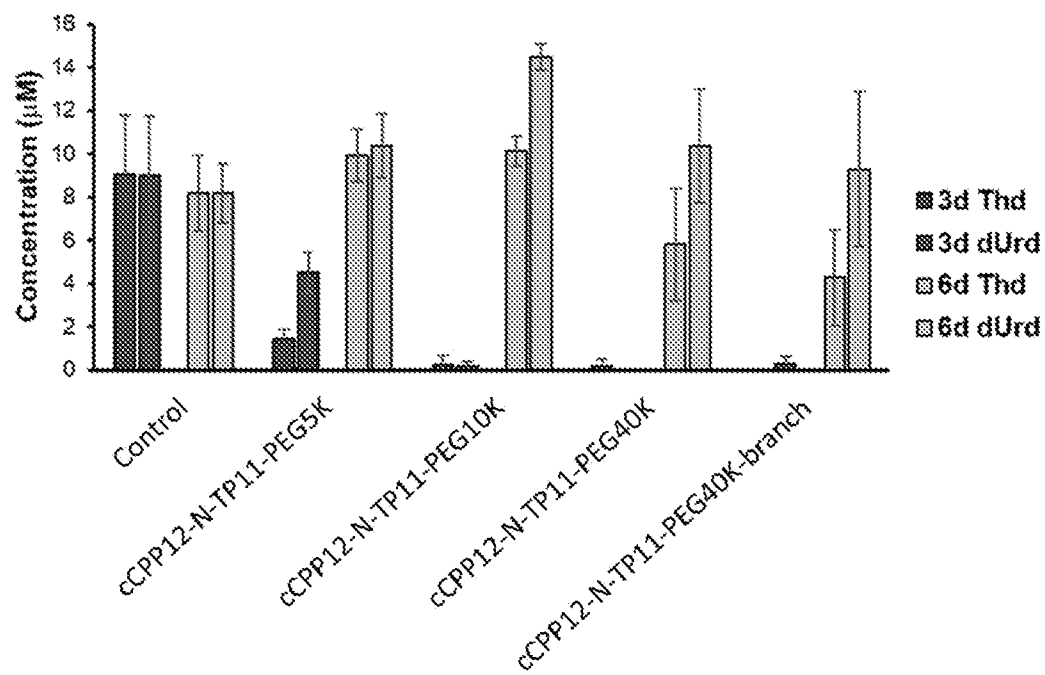


FIG.24A

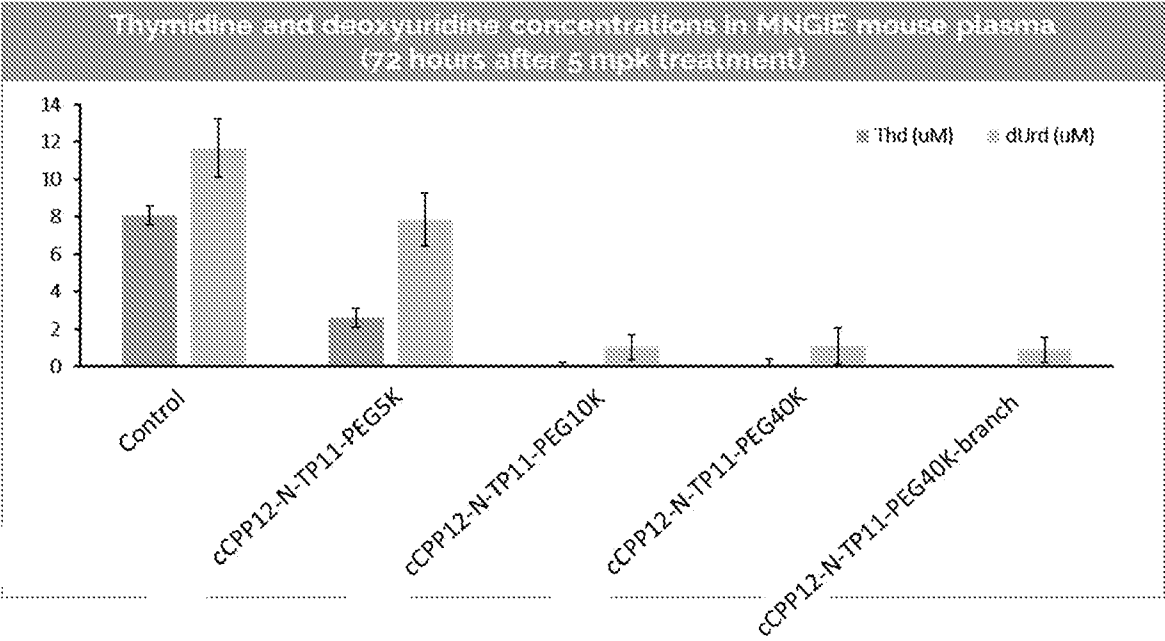


FIG. 24B

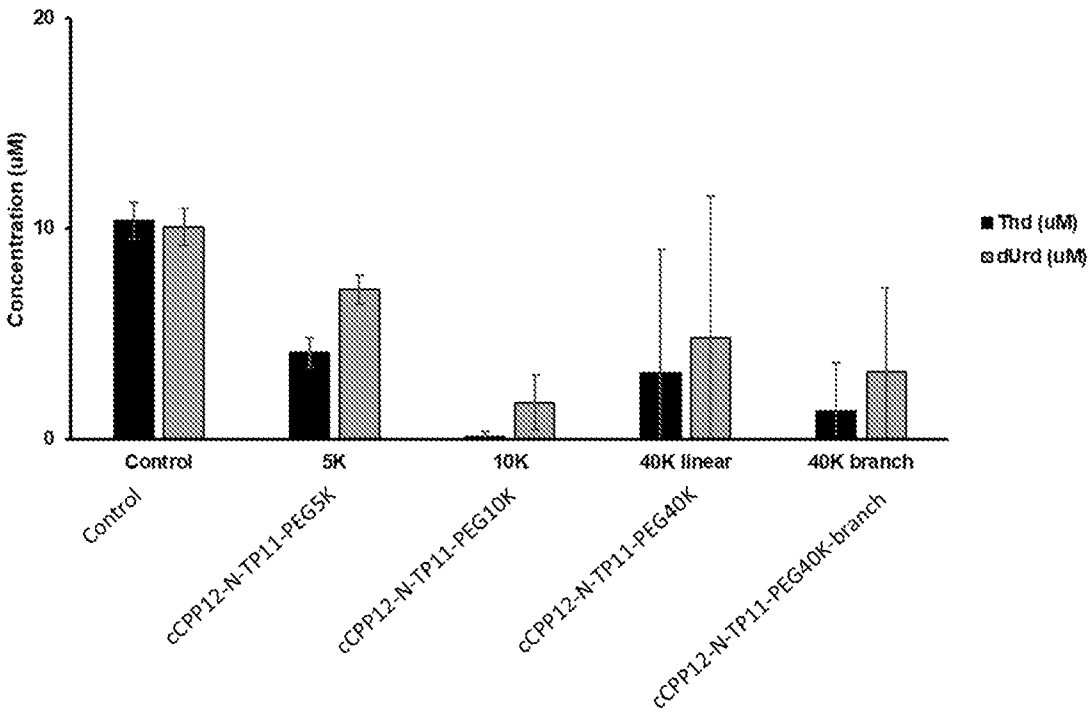
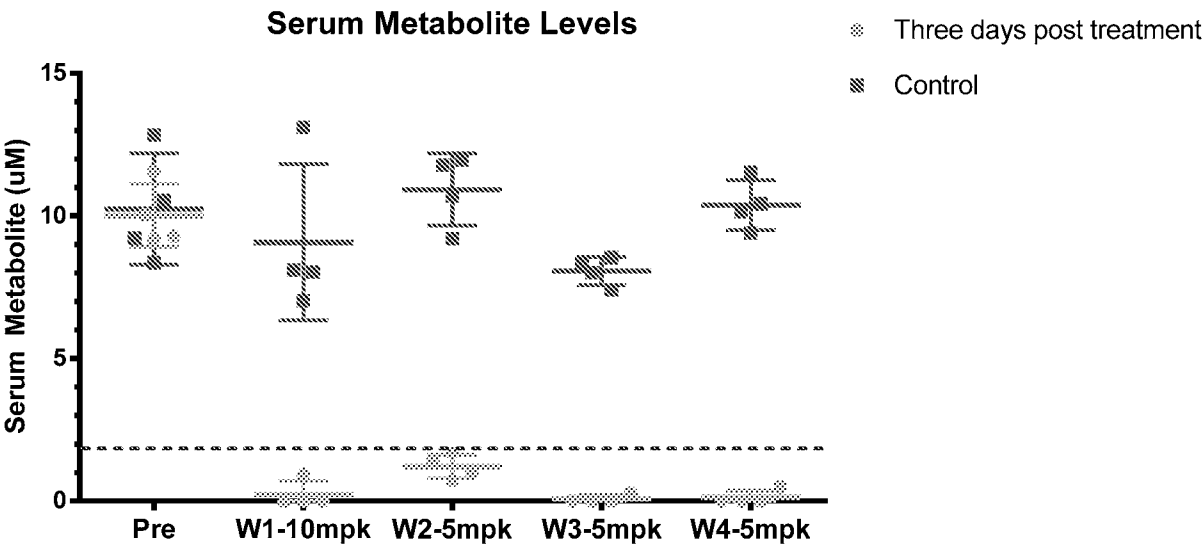


FIG. 25



20/27

FIG. 26A

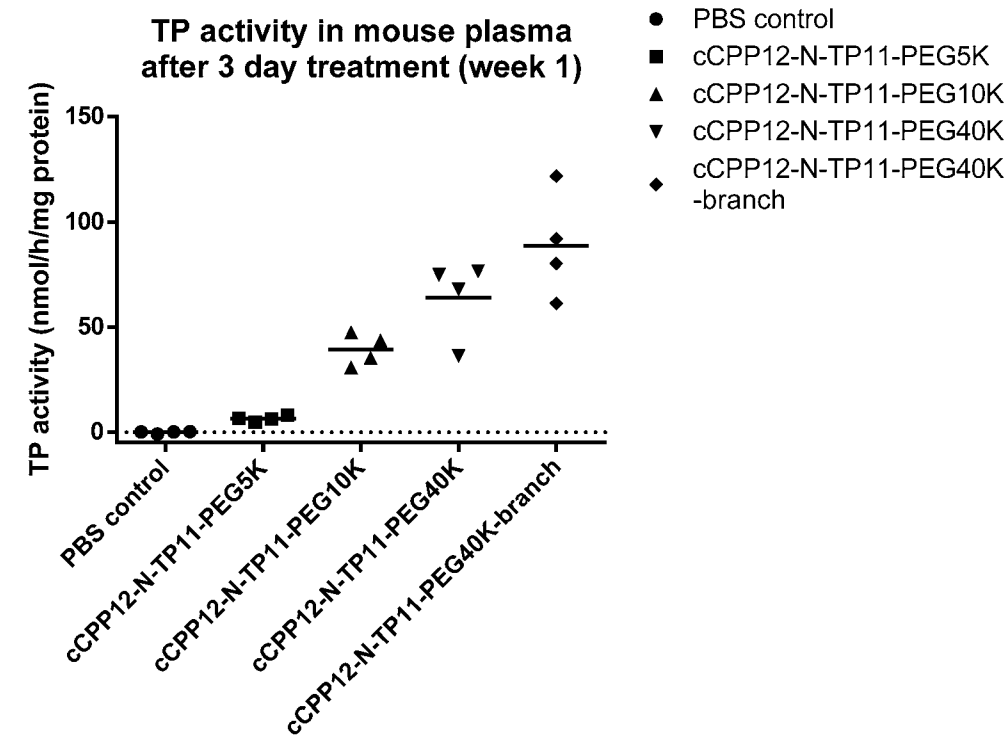


FIG. 26B

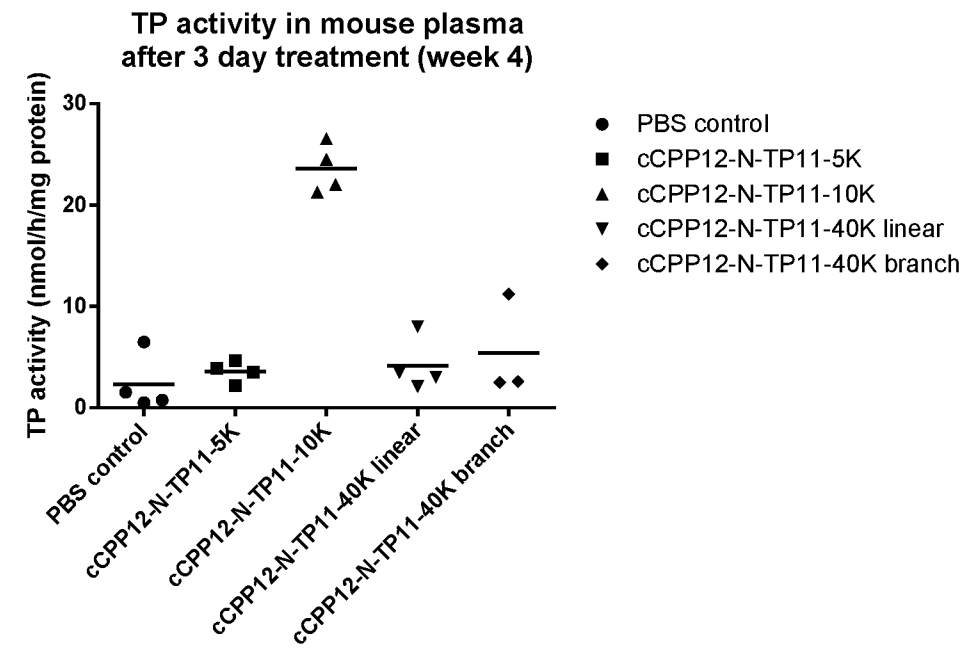


FIG. 27A

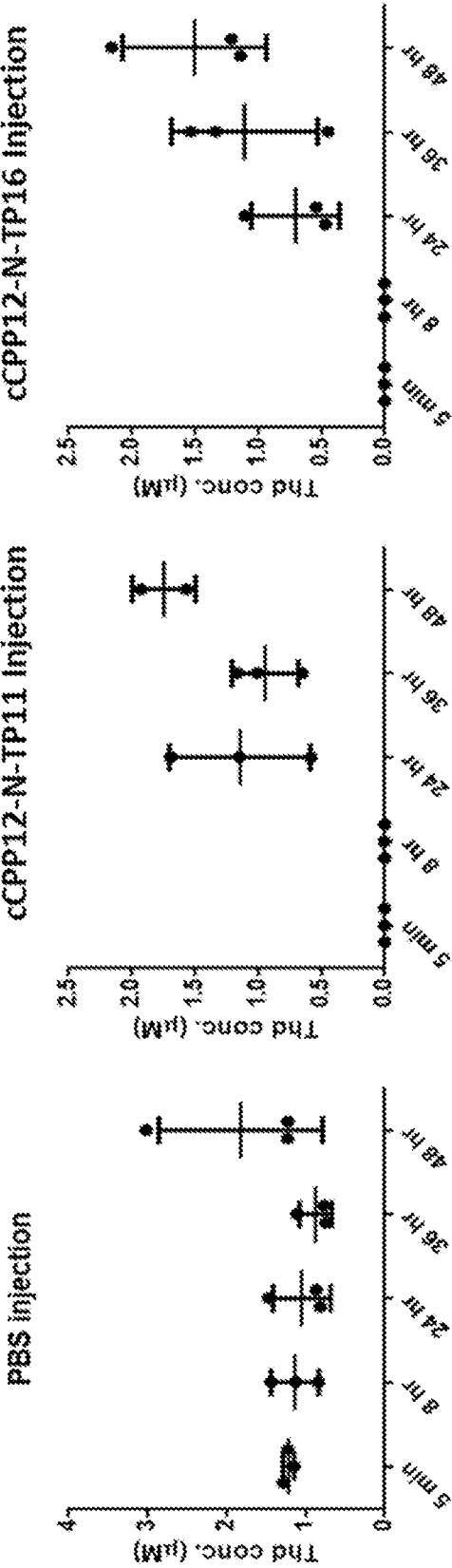


FIG. 27B

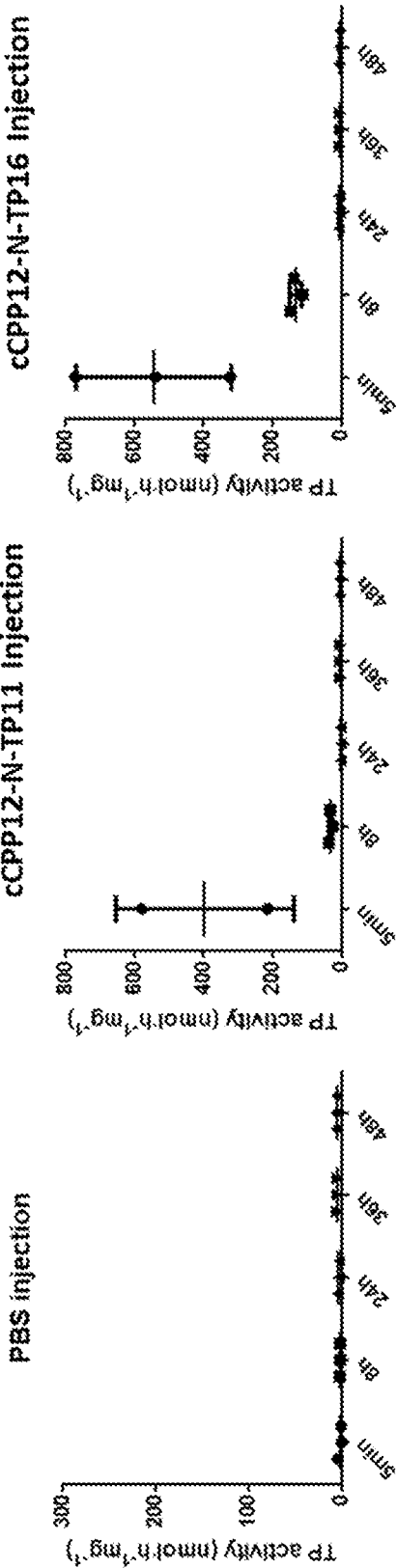


FIG. 28A

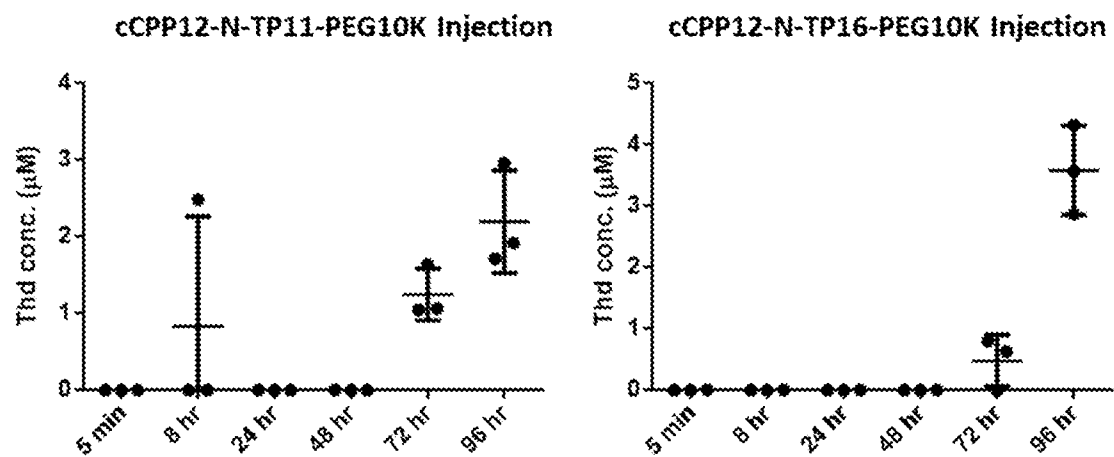
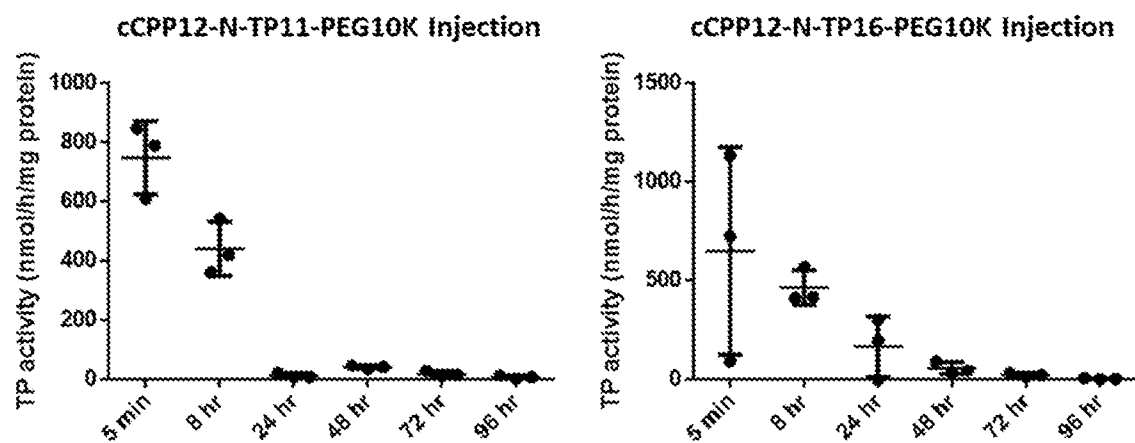


FIG.28B



23/27

FIG.29A

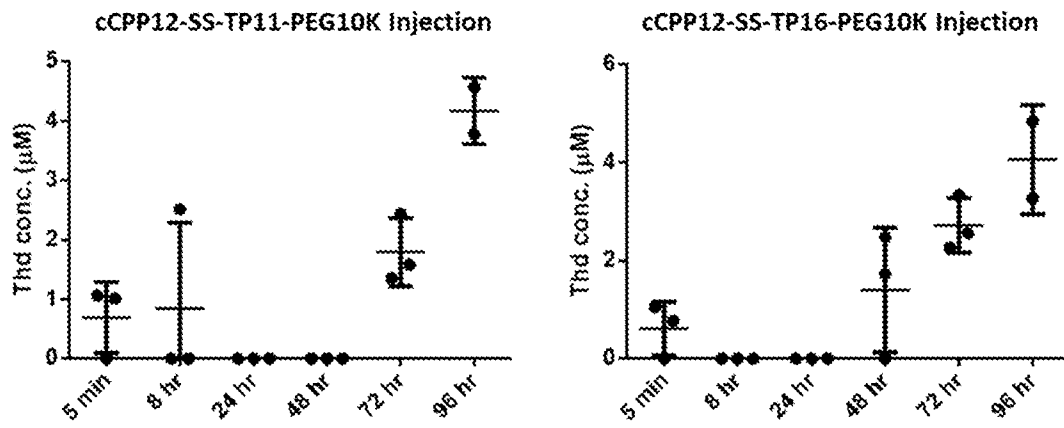


FIG.29B

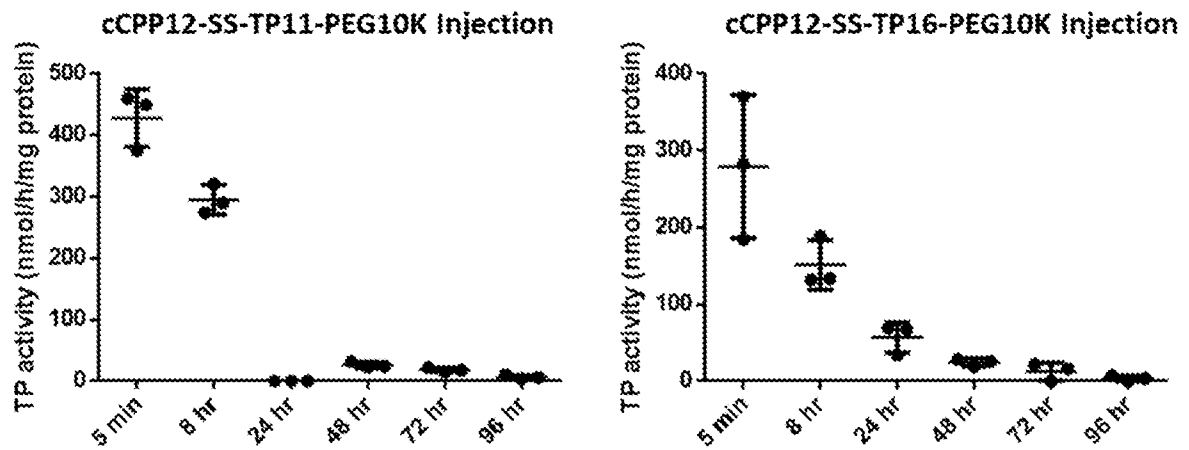


FIG. 30

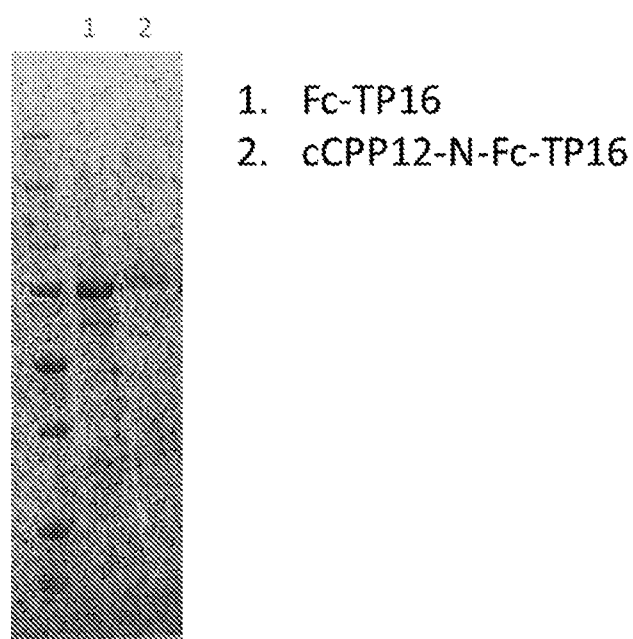


FIG. 31

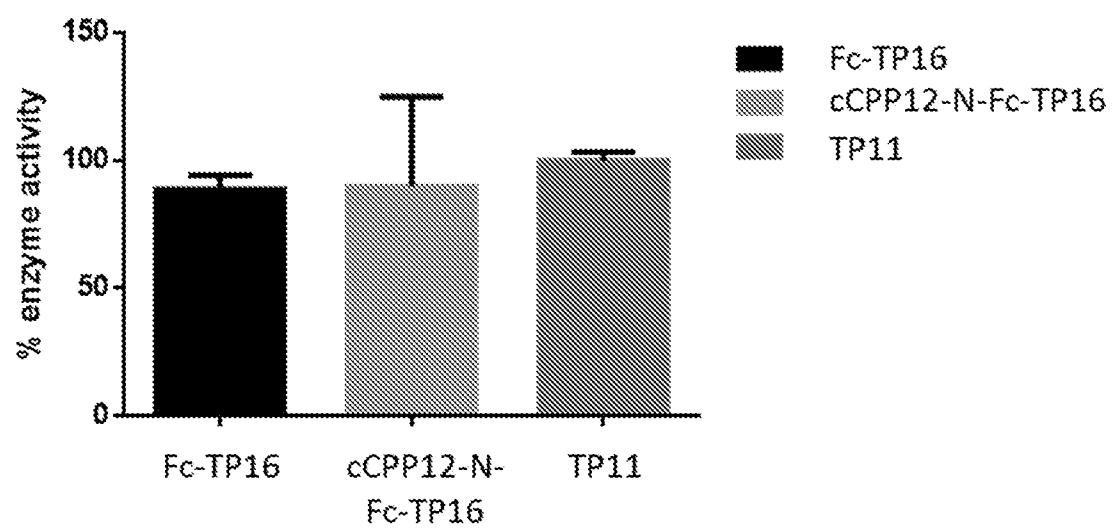


FIG. 32

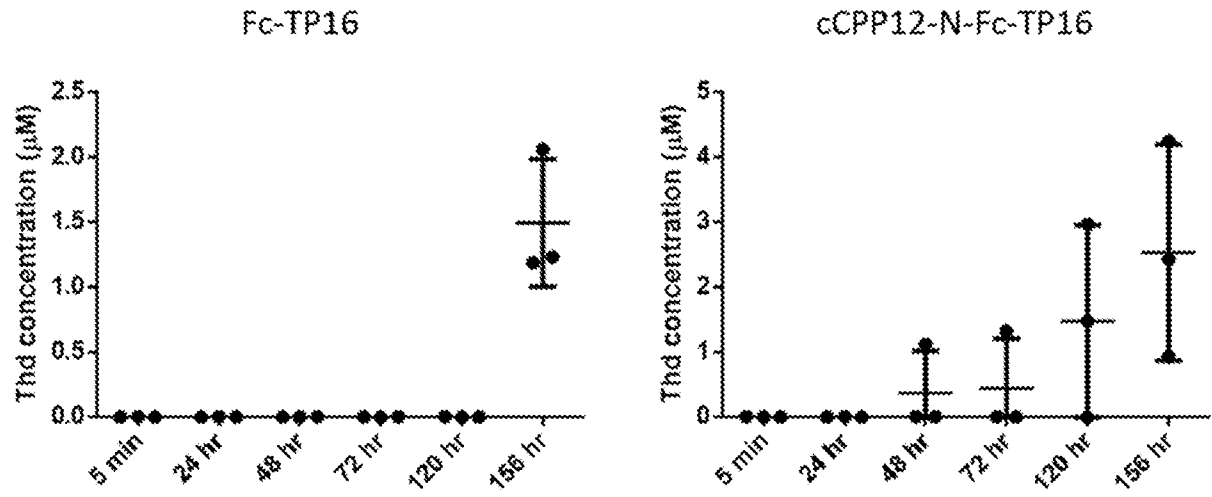


FIG. 33

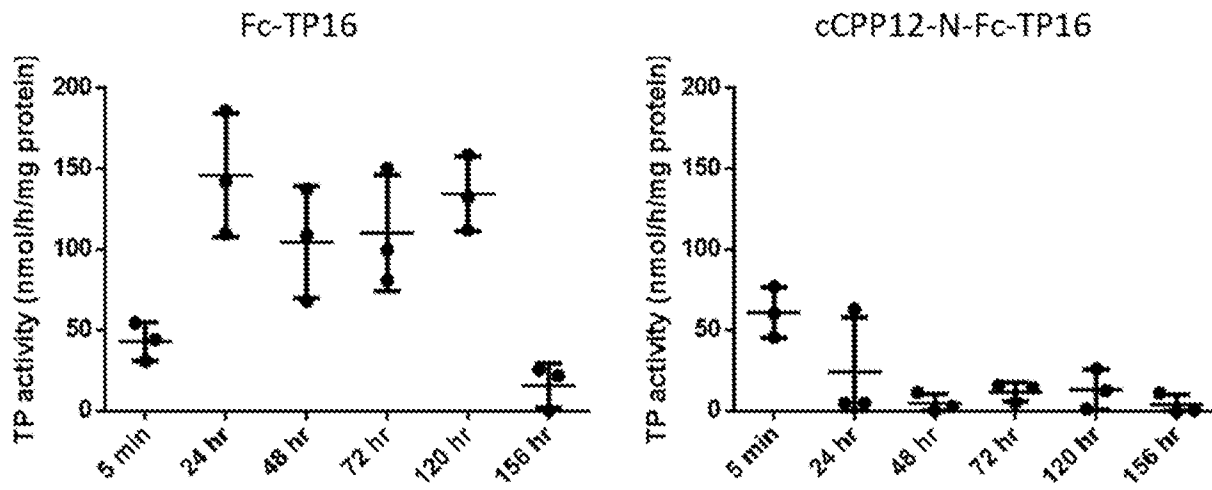


FIG. 34

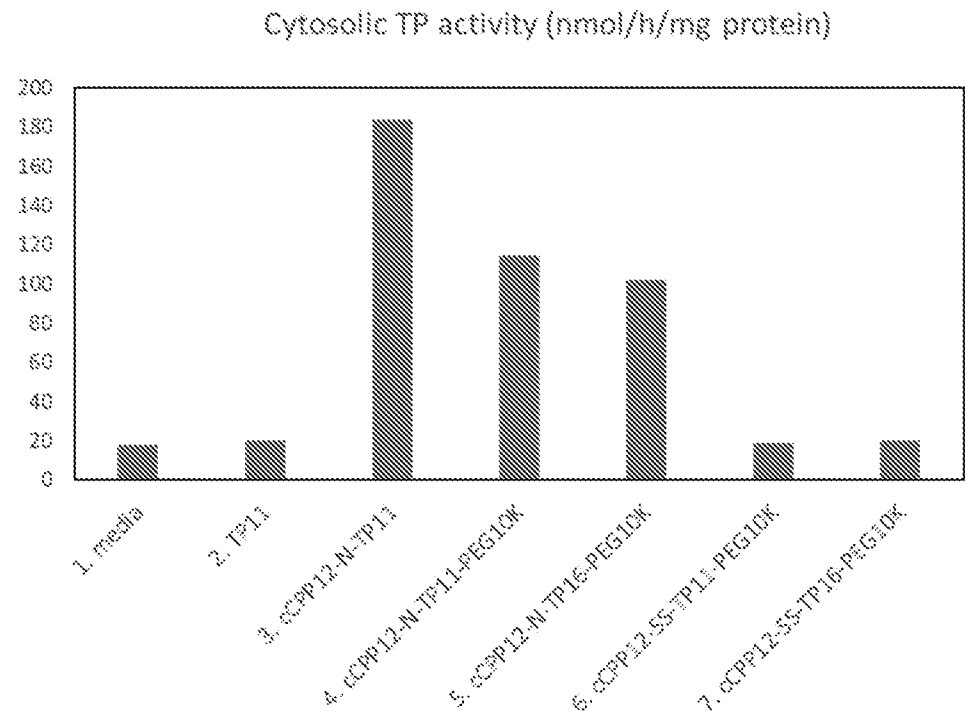
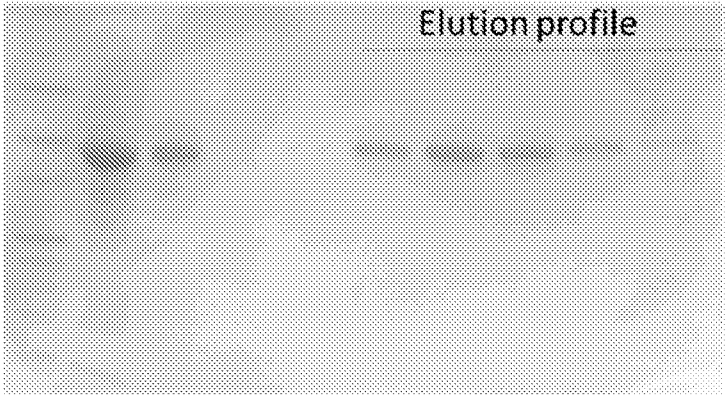
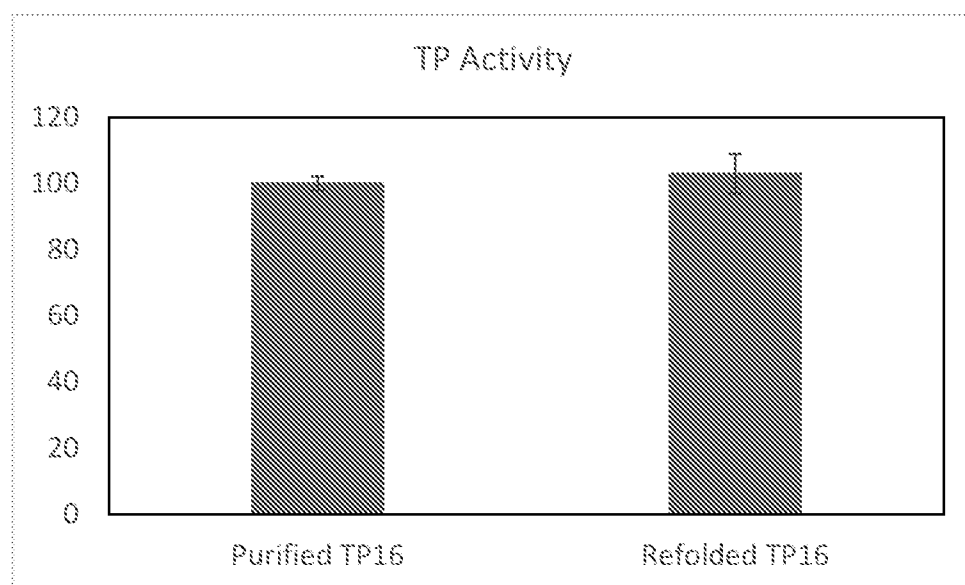


FIG. 35



- L: ladder
1. Dissolved inclusion body
 2. Refolded protein after rapid dilution

Elution profile after Q sepharase

FIG. 36

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US19/19117

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
- a. ☐ forming part of the international application as filed:
☐ in the form of an Annex C/ST.25 text file.
☐ on paper or in the form of an image file.
- b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
- c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

An Invitation to Furnish Nucleotide and/or Amino Acid Sequence Listing and to Pay, Where Applicable, Late Furnishing Fee ("ISA/225") was mailed on 01 March 2019 (01.03.2019). The applicant did not, within the prescribed time limit, pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13ter.1(a) or (b). Accordingly, ISA/US cannot consider the sequence listing submitted on 29 March 2019 (29.03.2019). Therefore, the international search has been carried out only to the extent possible

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US19/19117

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 10-44
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US19/19117

A. CLASSIFICATION OF SUBJECT MATTER

IPC - A61K 35/28, 48/00; C12N 5/078 (2019.01)

CPC - A61K 35/28, 48/00; C12N 5/06, 5/078 5/0641, 5/0644

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2018/0030411 A1 (RUBIUS THERAPEUTICS INC.) 01 February 2018; paragraphs [0611], [0630], [0659]	1-4, 5/2-4, 6/5/2-4, 7/5/2-4, 8/5/2-4, 9/5/2-4
A	US 2017/0020926 A1 (RUBIUS THERAPEUTICS INC.) 26 January 2017; entire document	1-4, 5/2-4, 6/5/2-4, 7/5/2-4, 8/5/2-4, 9/5/2-4
A	US 2014/0235557 A1 (DE WAARD) 21 August 2014; entire document	1-4, 5/2-4, 6/5/2-4, 7/5/2-4, 8/5/2-4, 9/5/2-4

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 April 2019 (30.05.2019)

Date of mailing of the international search report

13 MAY 2019

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774