



US 20250122242A1

(19) **United States**

(12) **Patent Application Publication**
Wang et al.

(10) **Pub. No.: US 2025/0122242 A1**

(43) **Pub. Date: Apr. 17, 2025**

(54) **SELECTION SYSTEMS, PEPTIDES
DETERMINED THEREWITH, AND
METHODS OF USING SAME**

Related U.S. Application Data

(60) Provisional application No. 63/590,069, filed on Oct. 13, 2023.

(71) Applicant: **Wisconsin Alumni Research
Foundation**, Madison, WI (US)

Publication Classification

(51) **Int. Cl.**
C07K 7/64 (2006.01)

(52) **U.S. Cl.**
CPC **C07K 7/64** (2013.01)

(73) Assignee: **Wisconsin Alumni Research
Foundation**, Madison, WI (US)

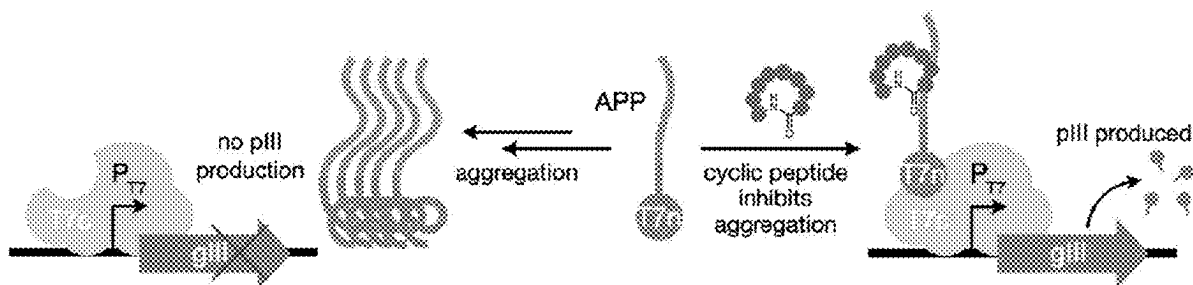
(57) **ABSTRACT**

Selection systems, such as selection systems for determining peptides that inhibit protein aggregation, peptides determined with the selection systems, and methods of using the selection systems and the peptides.

Specification includes a Sequence Listing.

(21) Appl. No.: **18/906,707**

(22) Filed: **Oct. 4, 2024**



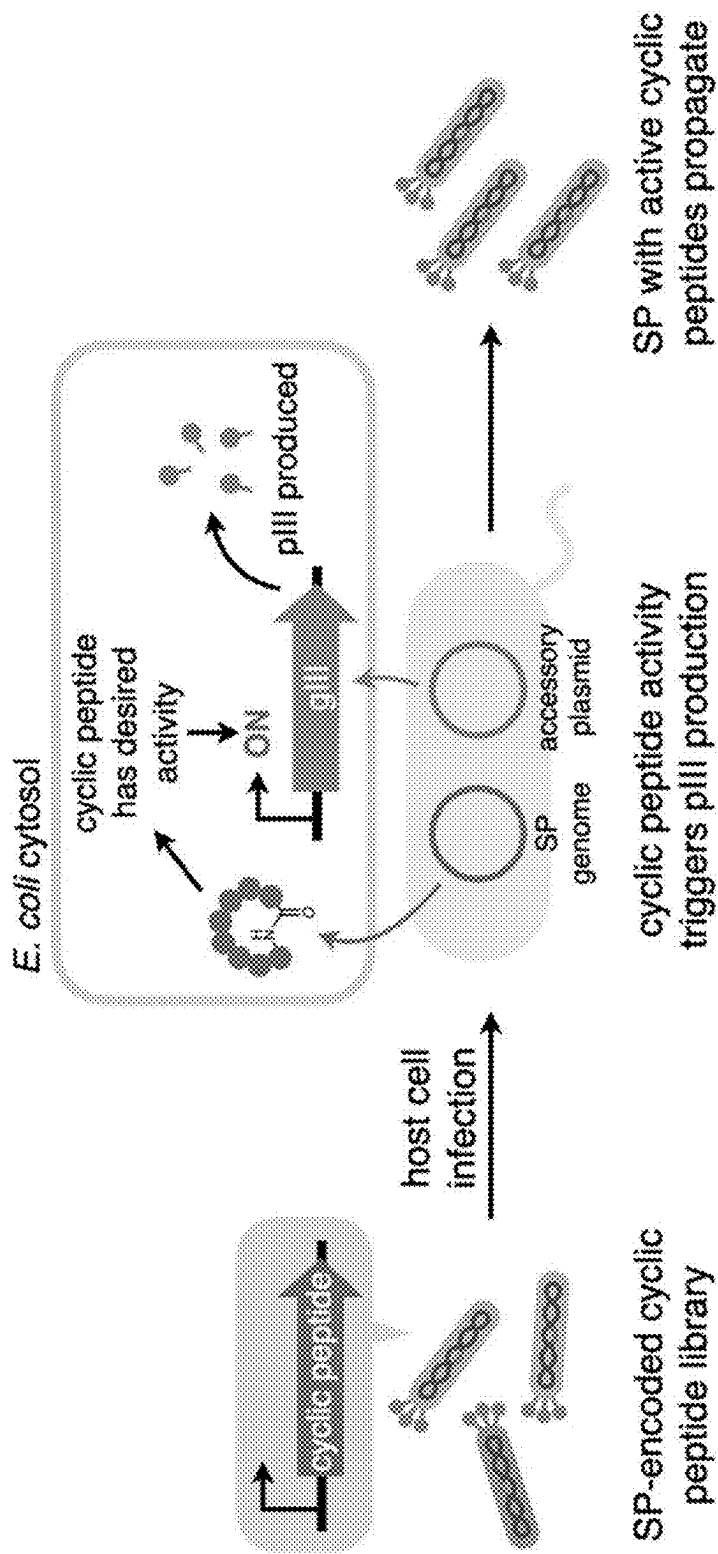


FIG. 1

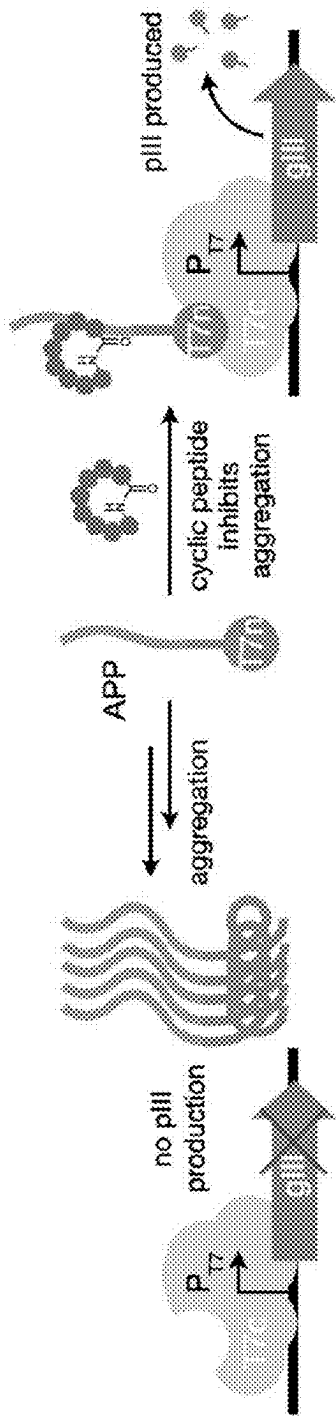


FIG. 2A

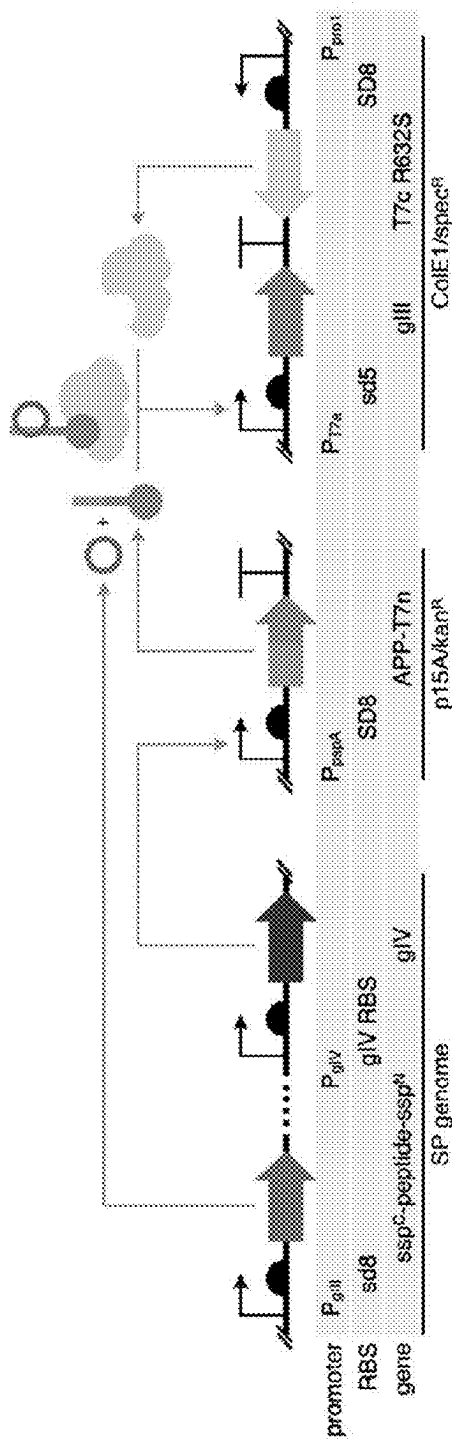


FIG. 2B

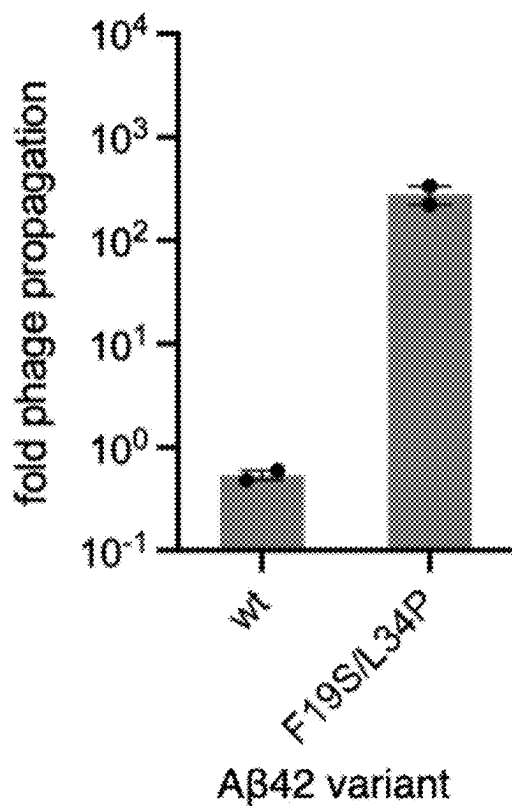


FIG. 2C

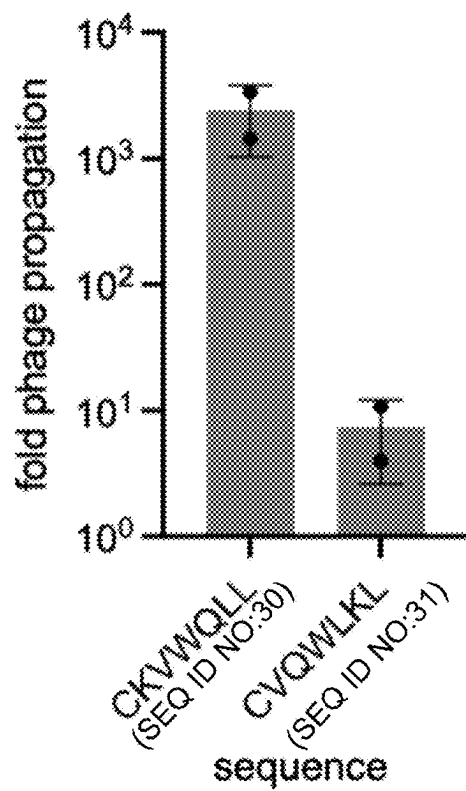


FIG. 2D

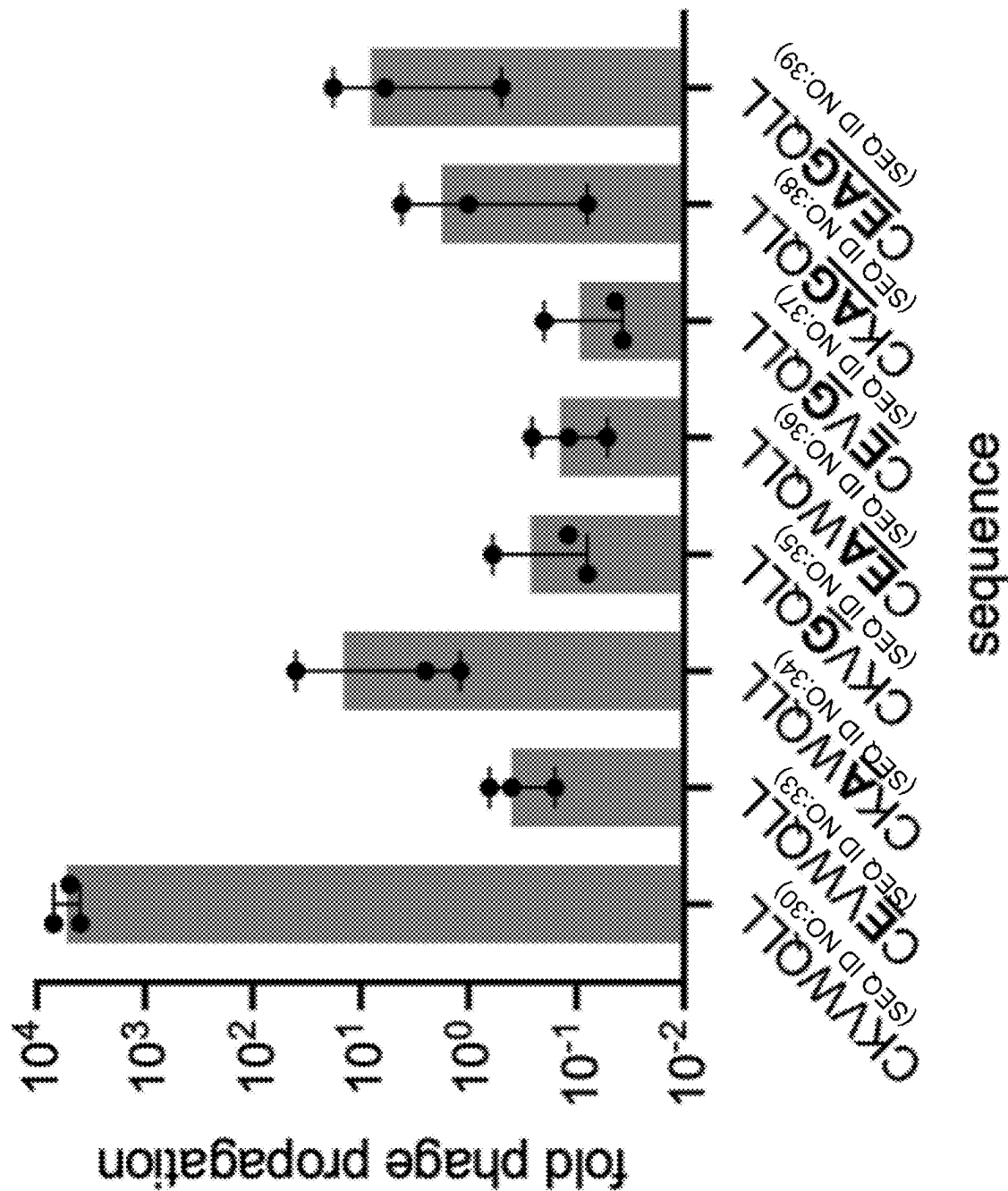


FIG. 2E

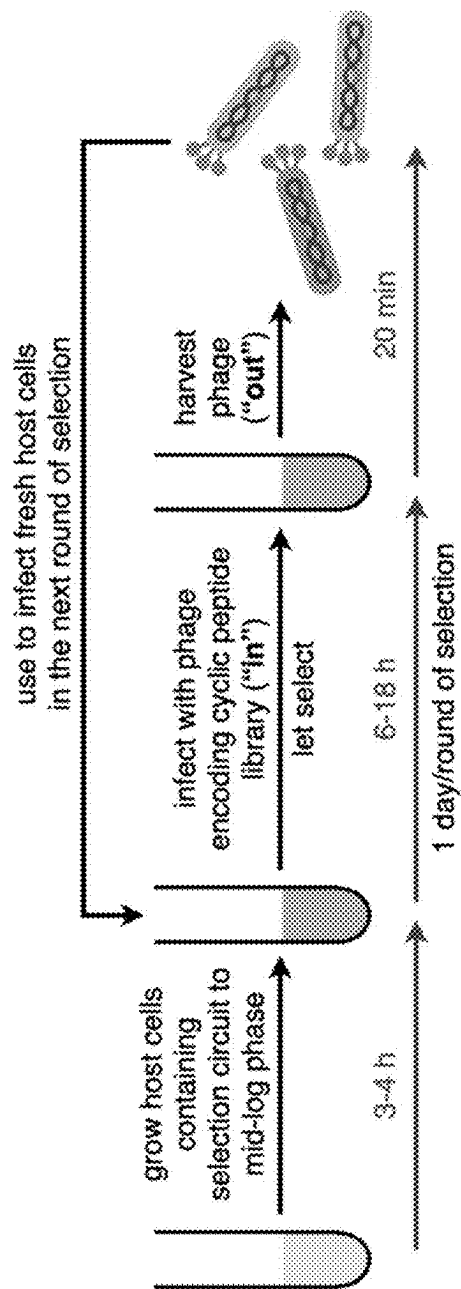


FIG. 3A

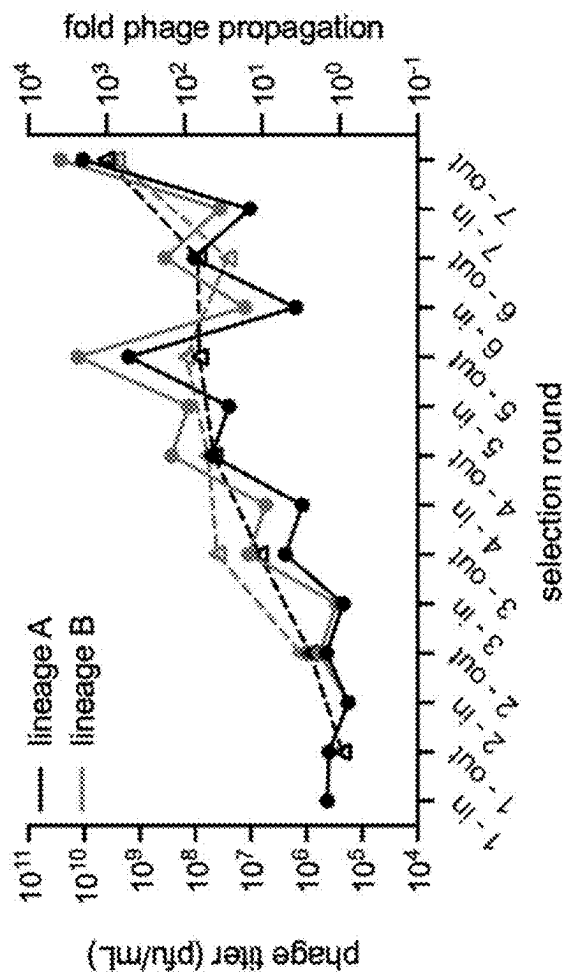


FIG. 3B

	CP sequence							
library	C	X	X	X	Q	L	X	
A-1	C	R	V	W	Q	L	C	SEQ ID NO:40
A-2	C	K	V	W	Q	L	L	SEQ ID NO:30
A-3	C	K	V	W	Q	L	L	SEQ ID NO:30
A-4	C	K	V	W	Q	L	L	SEQ ID NO:30
B-1	C	I	V	W	Q	L	C	SEQ ID NO:41
B-2	C	K	V	W	Q	L	L	SEQ ID NO:30
B-3	C	K	V	W	Q	L	L	SEQ ID NO:30
B-4	C	K	V	W	Q	L	A	SEQ ID NO:42

FIG. 3C

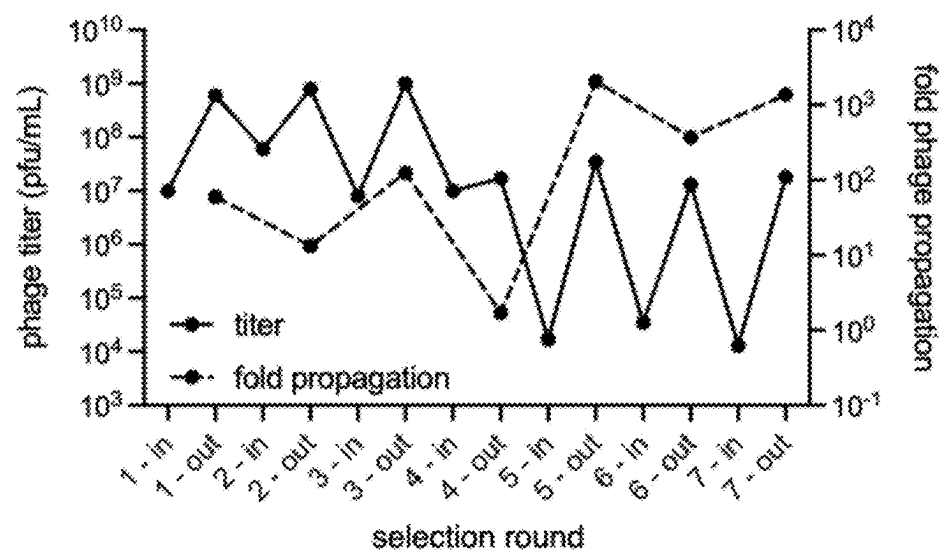


FIG. 3D

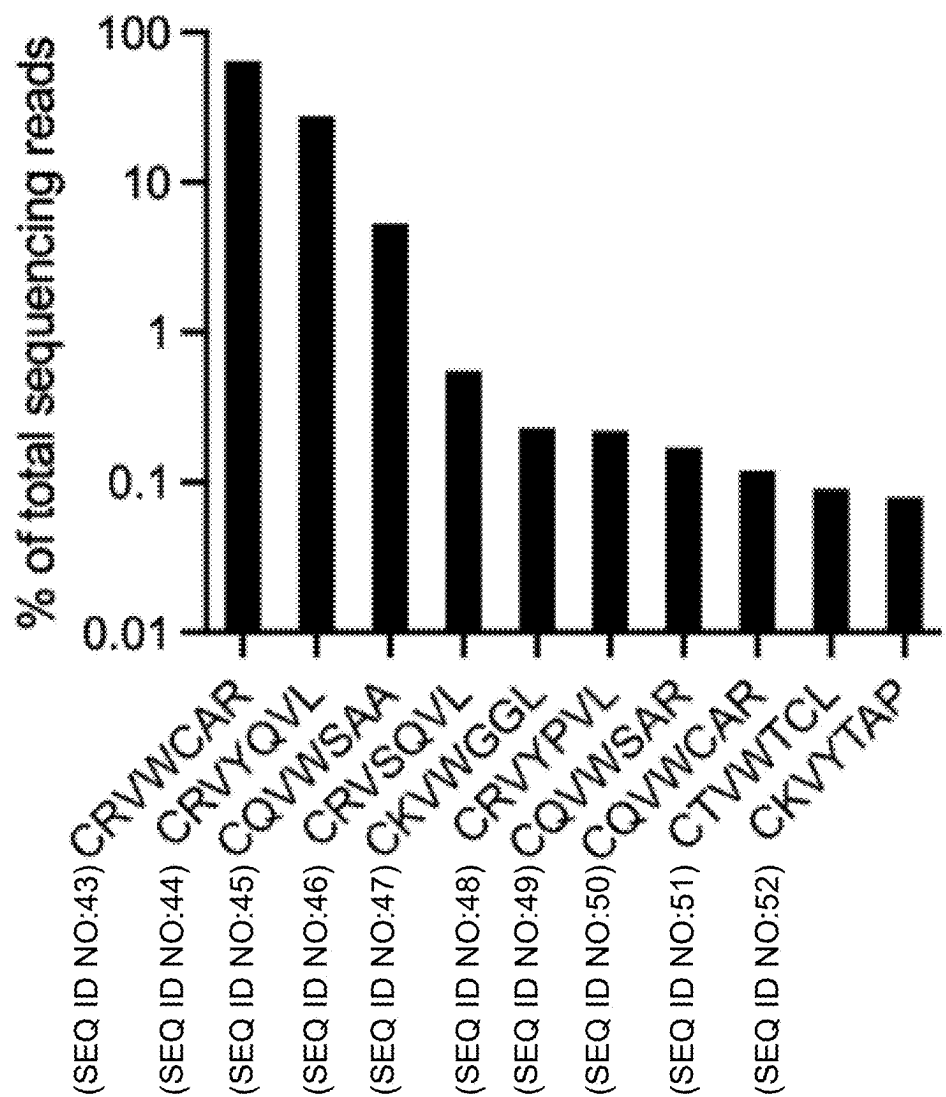


FIG. 3E

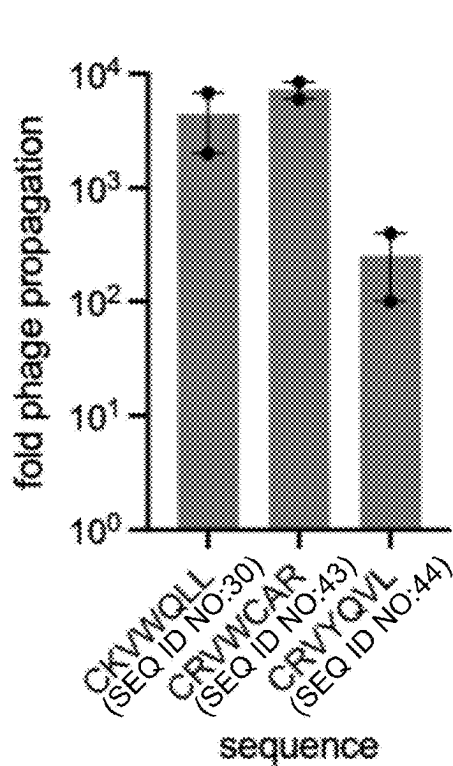


FIG. 3F

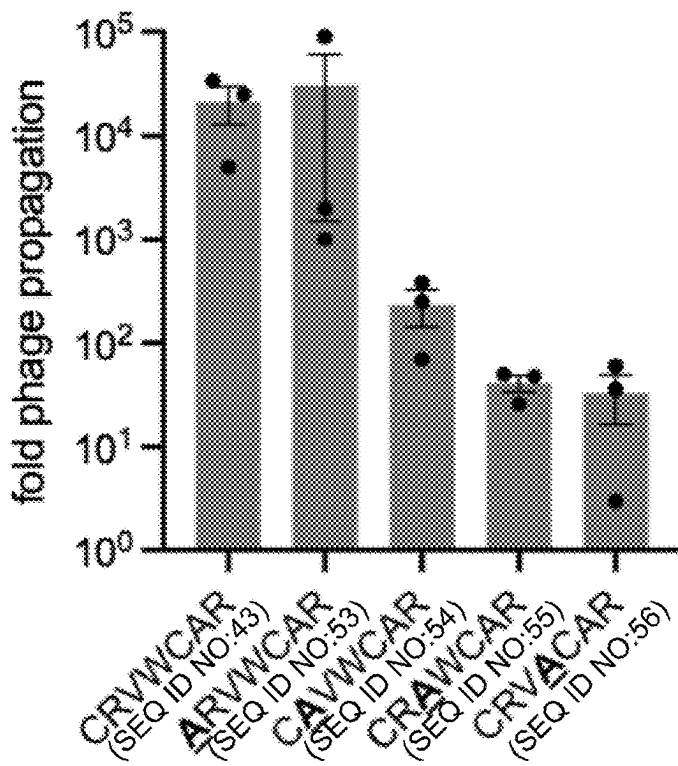


FIG. 3G

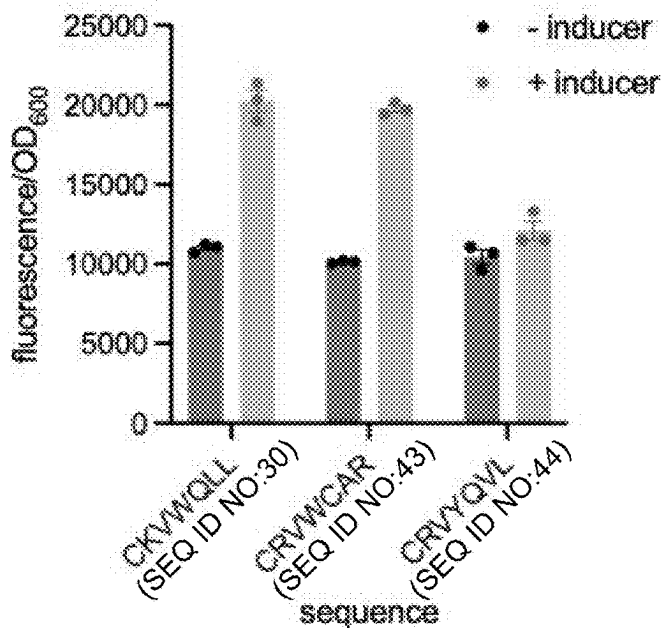


FIG. 3H

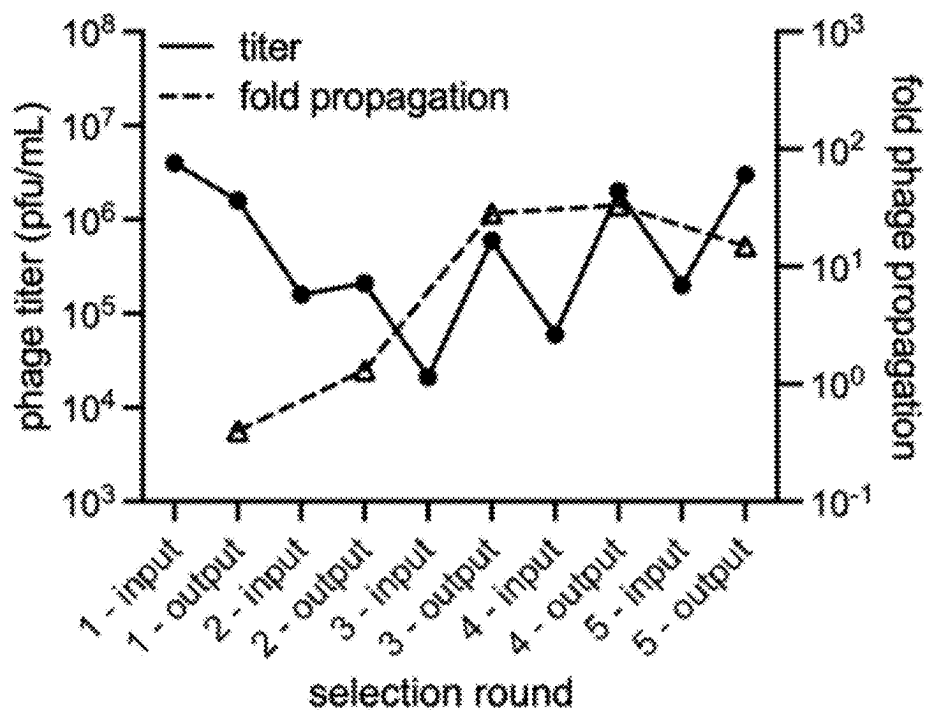


FIG. 4A

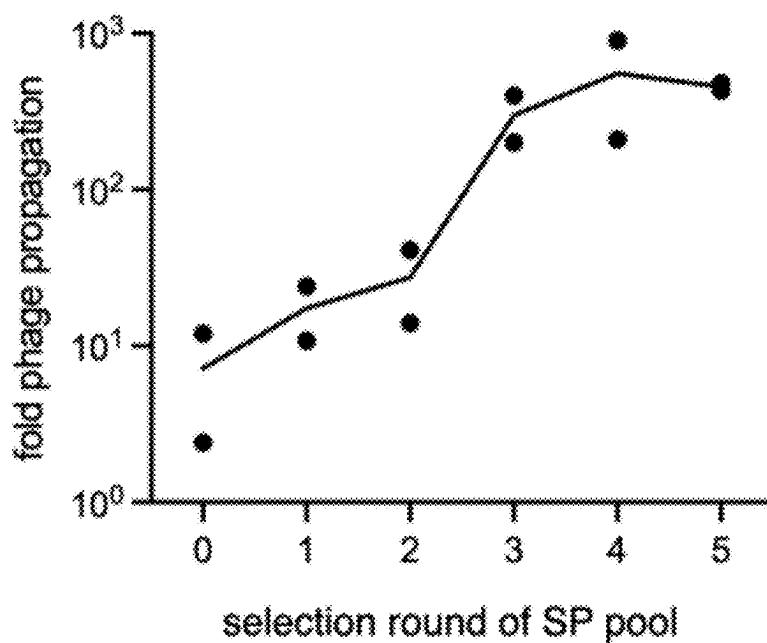


FIG. 4B

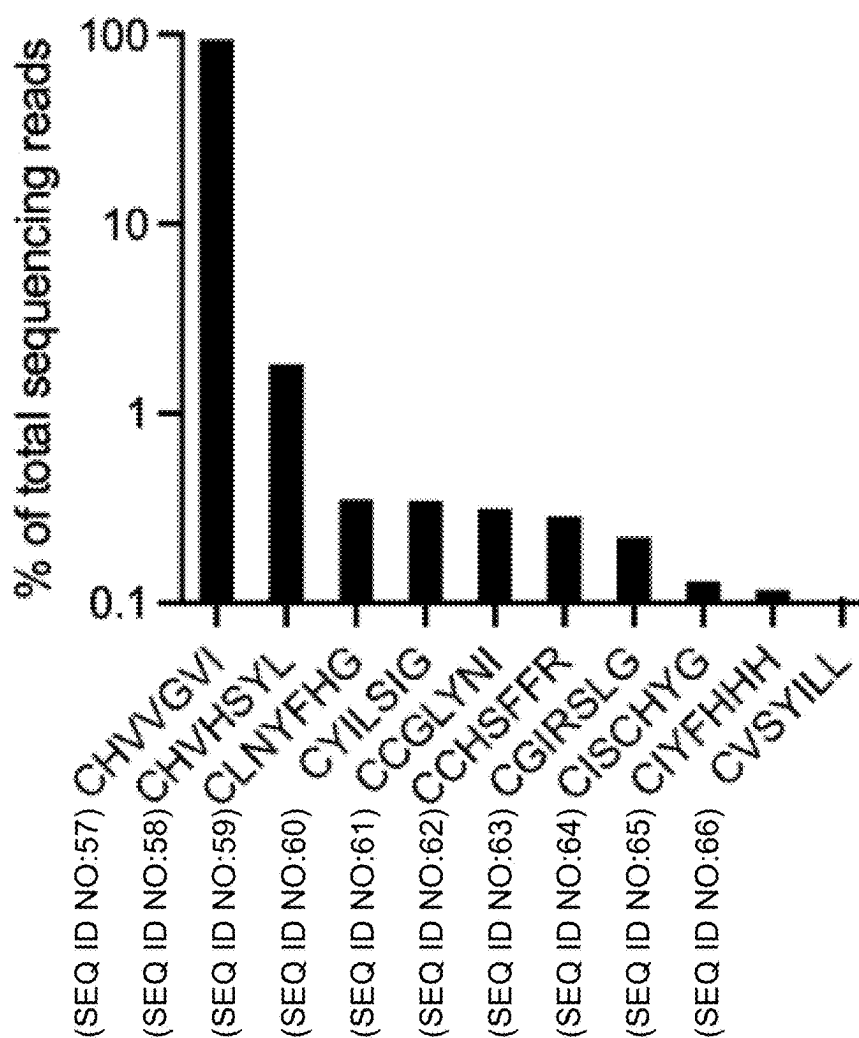
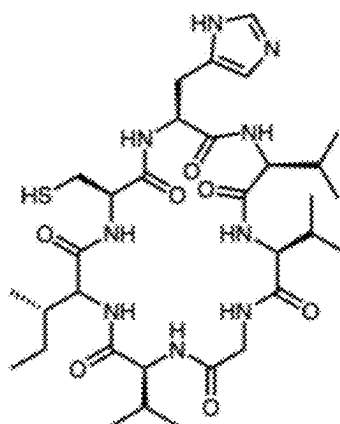


FIG. 4C



cyclo-CHVVGVI
(SEQ ID NO:57)

FIG. 4D

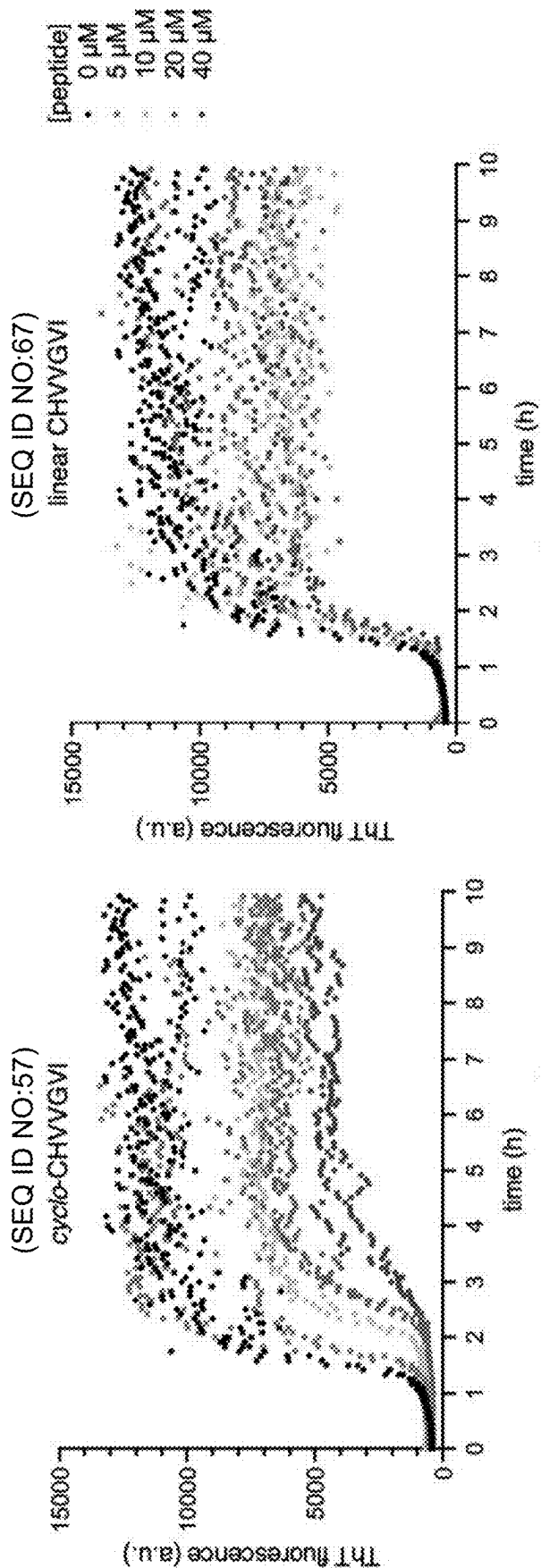


FIG. 4E

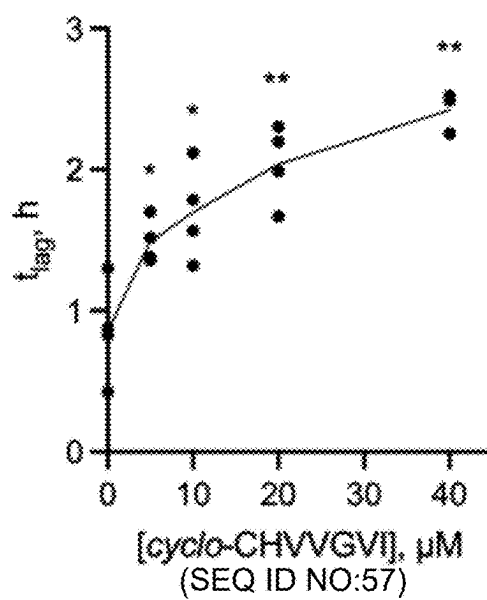


FIG. 4F

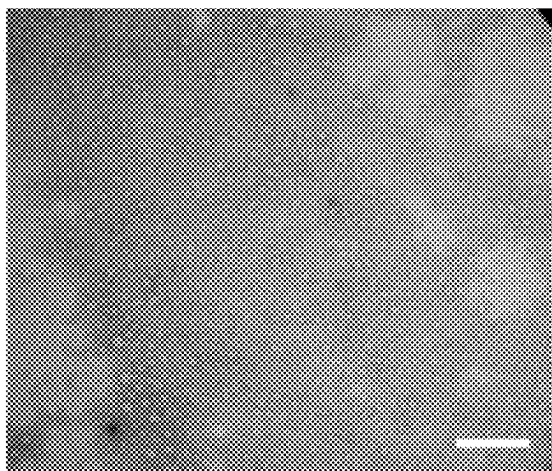


FIG. 4G

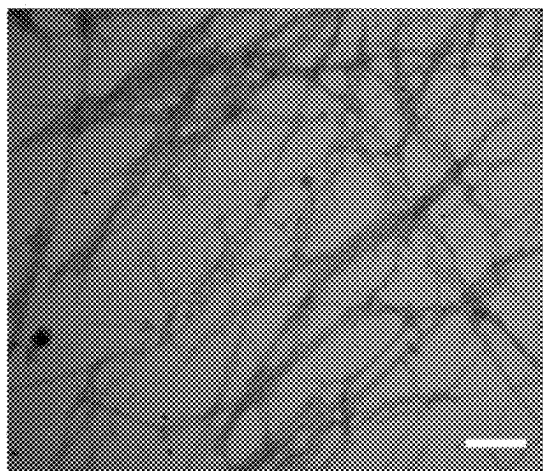


FIG. 4H

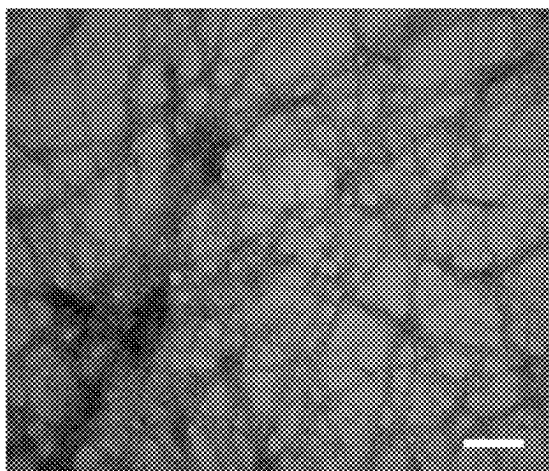


FIG. 4I

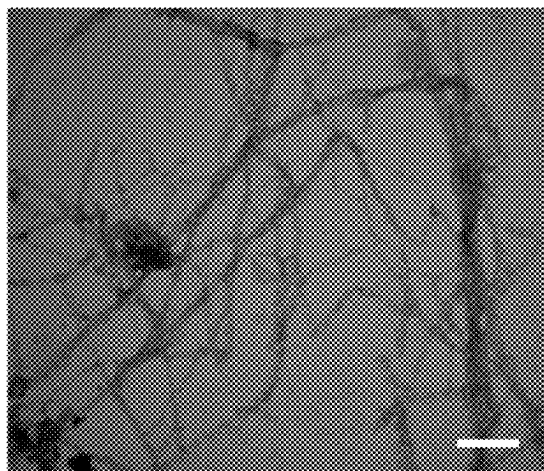


FIG. 4J

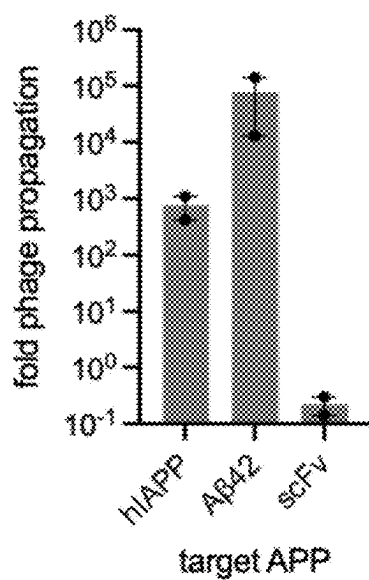


FIG. 5A

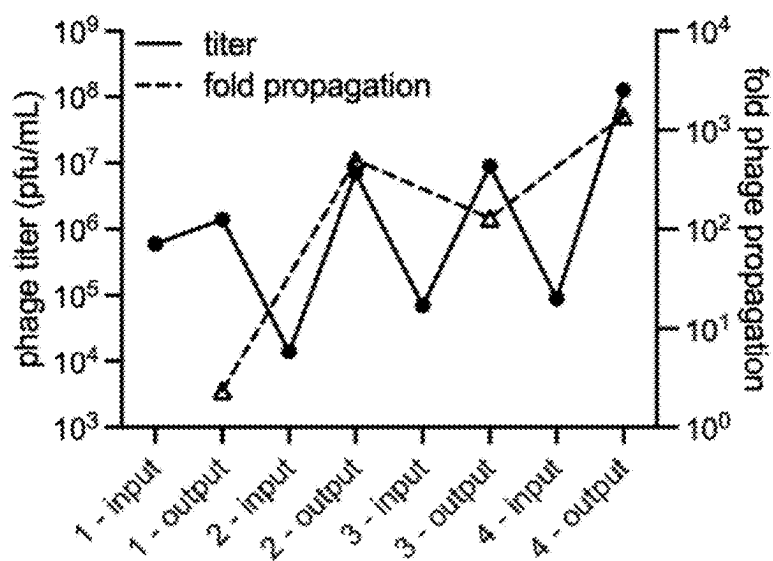


FIG. 5B

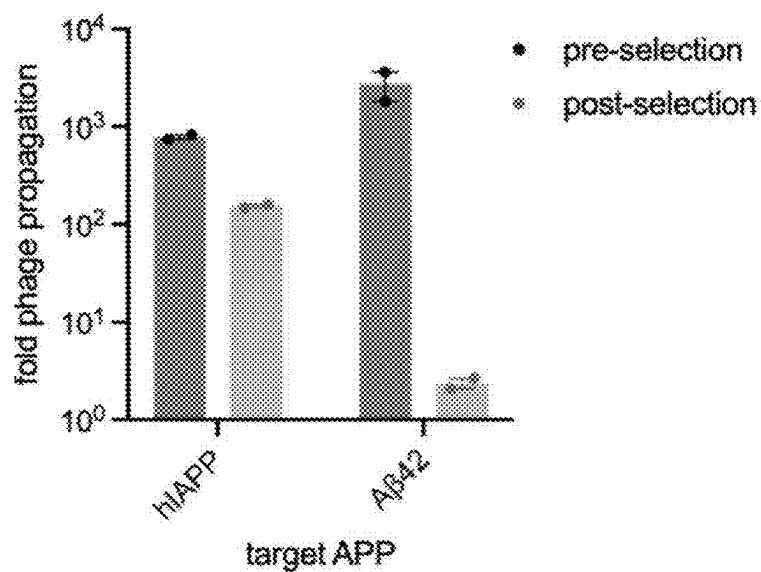


FIG. 5C

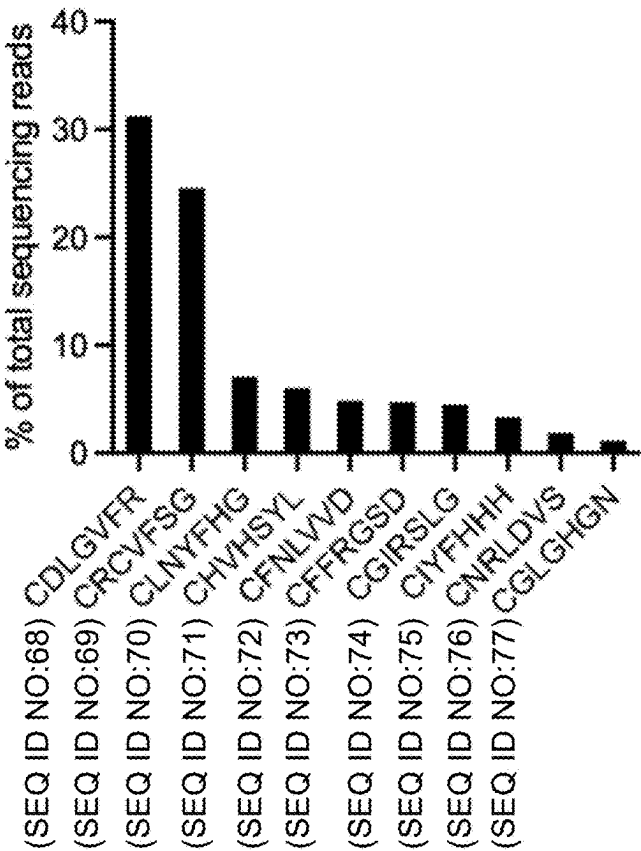


FIG. 5D

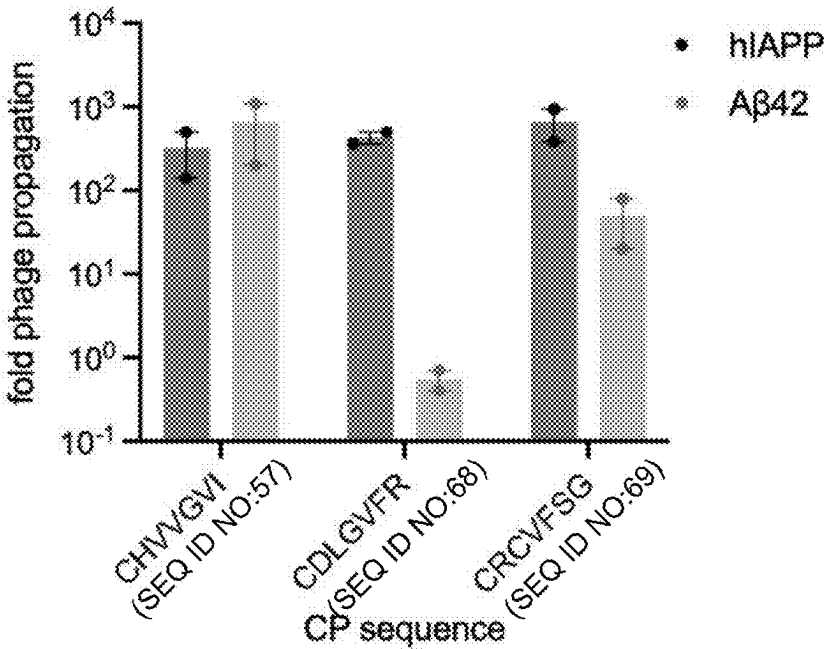
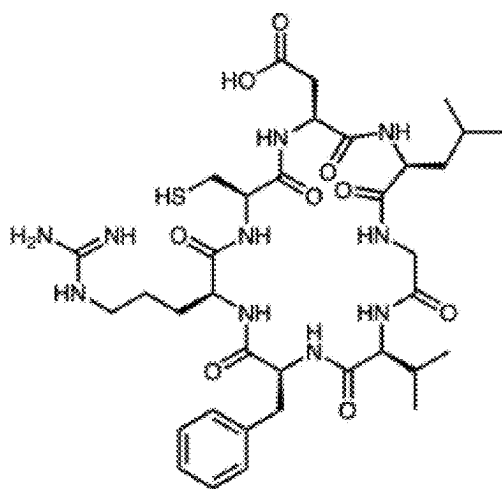
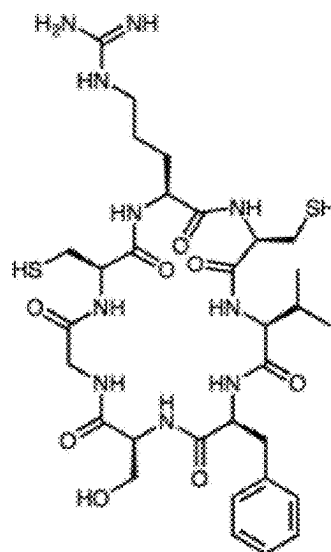


FIG. 5E



cyclo-CDLGVFR
(SEQ ID NO:68)

FIG. 6A



cyclo-CRCVFSG
(SEQ ID NO:69)

FIG. 6B

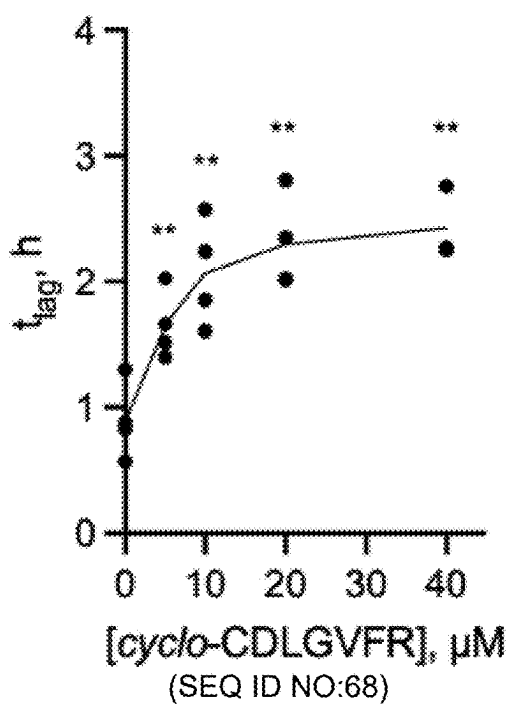


FIG. 6D

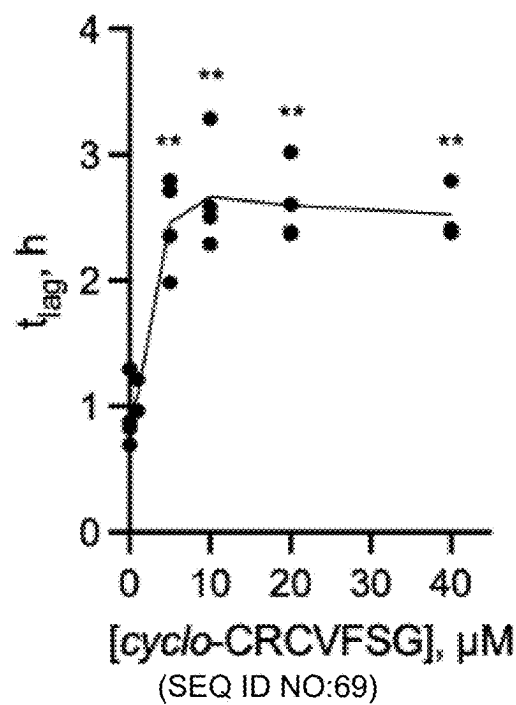


FIG. 6F

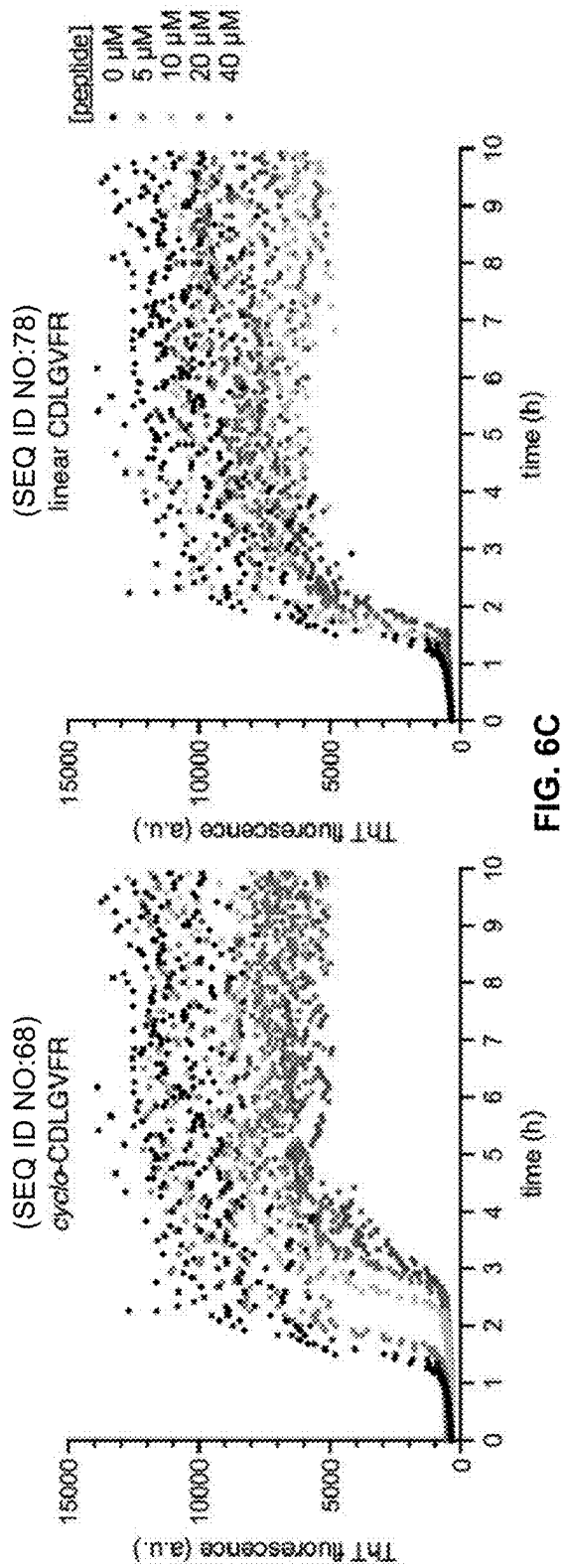


FIG. 6C

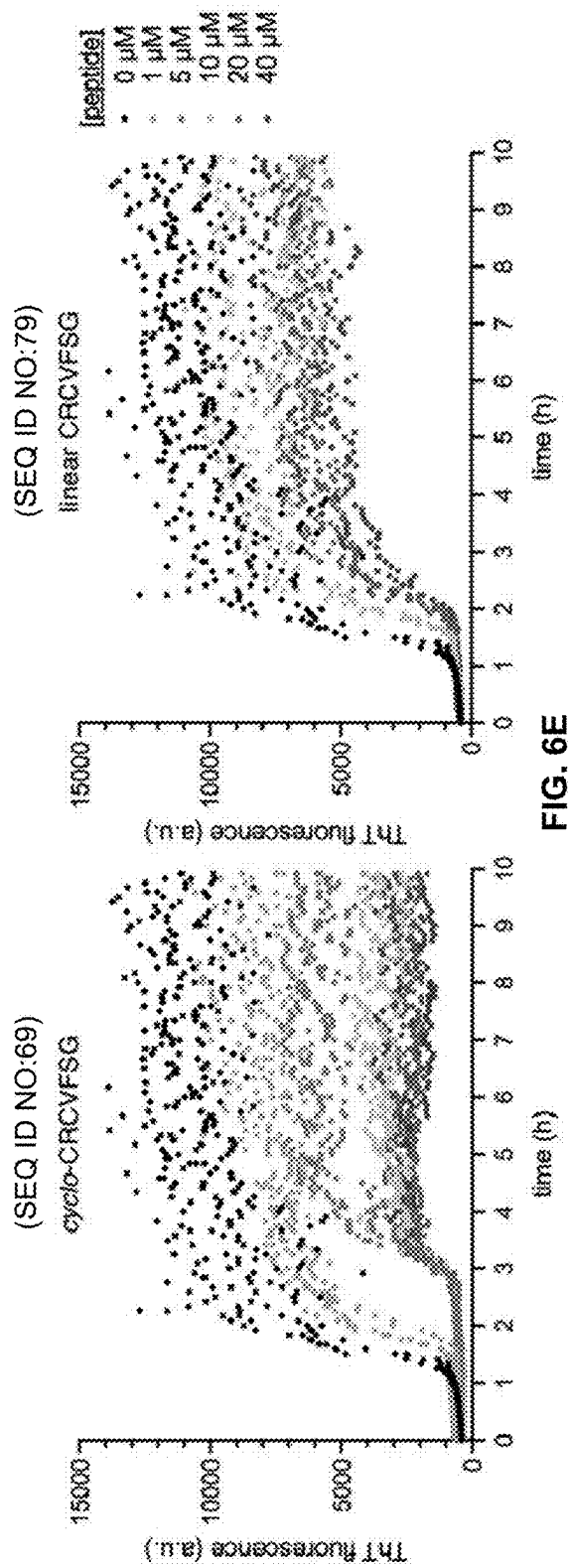


FIG. 6E

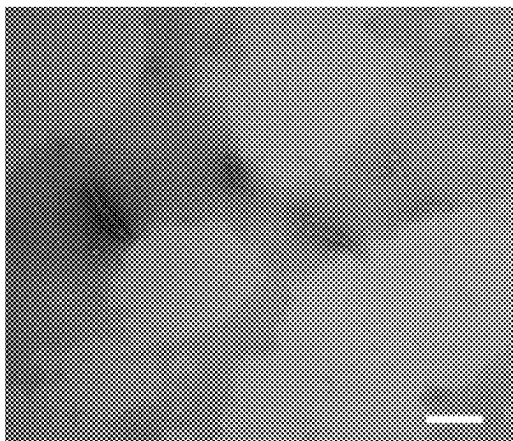


FIG. 6G

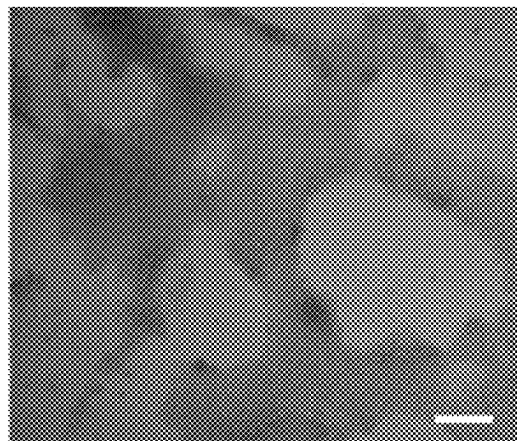


FIG. 6H

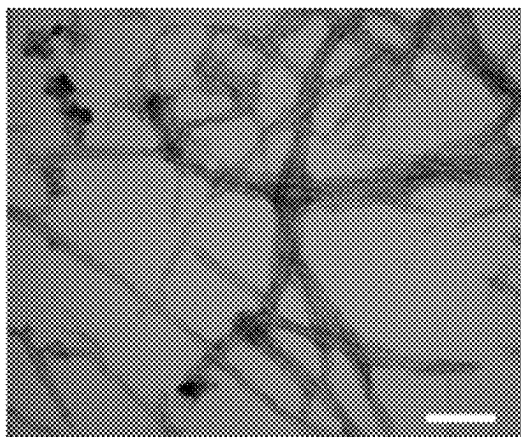


FIG. 6I

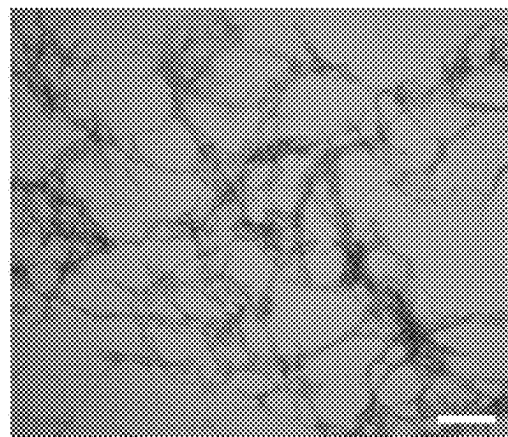


FIG. 6J

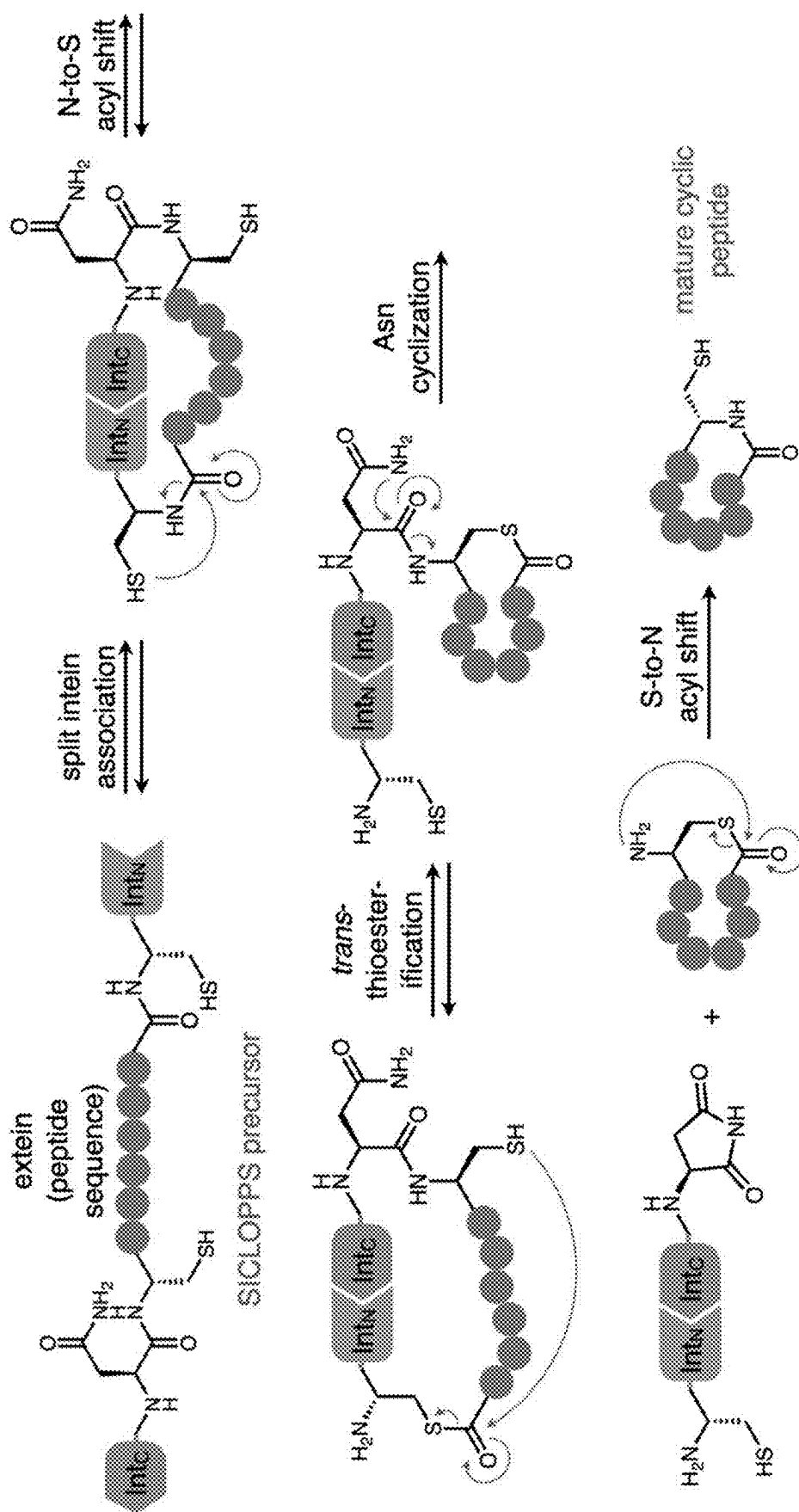
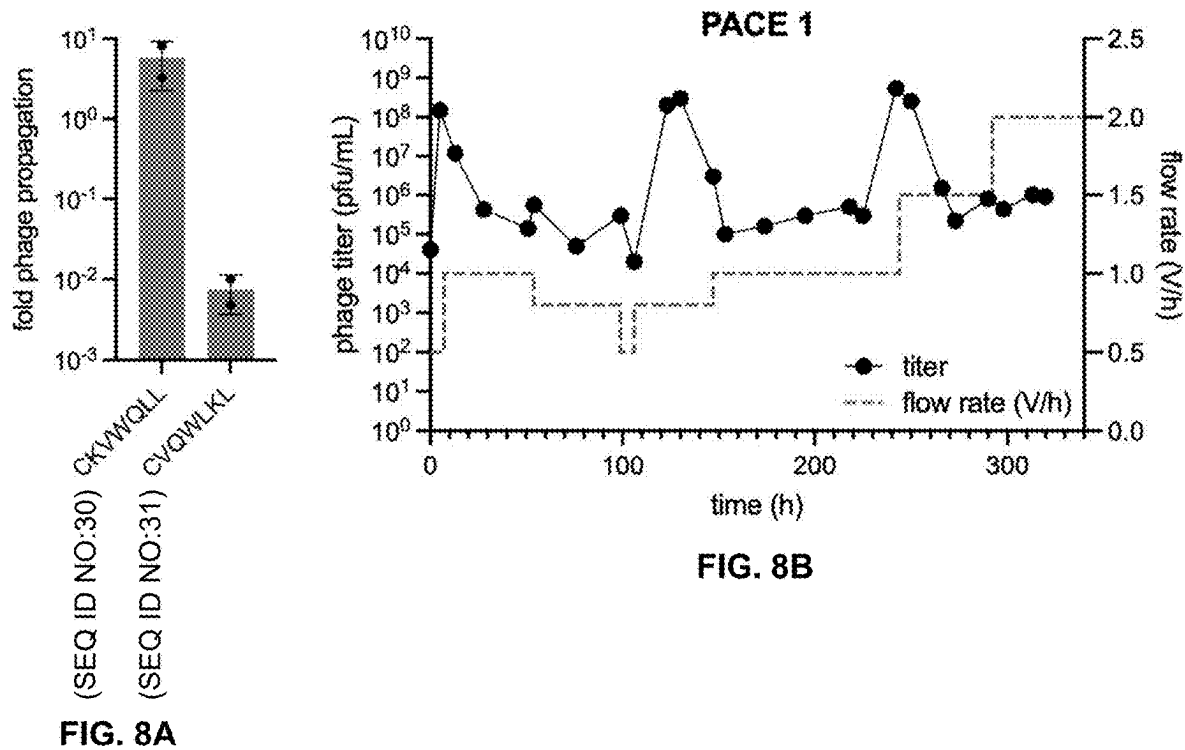


FIG. 7



	SspC	peptide	SspN						
AA position			20	36	44	98	104	105	119
parent			K	E	I	E	E	E	A
clone 1			N	K		EKN*			
clone 2				K		EKN*			
clone 3				K		EKN*			
clone 4				K		EKN*			

FIG. 8C

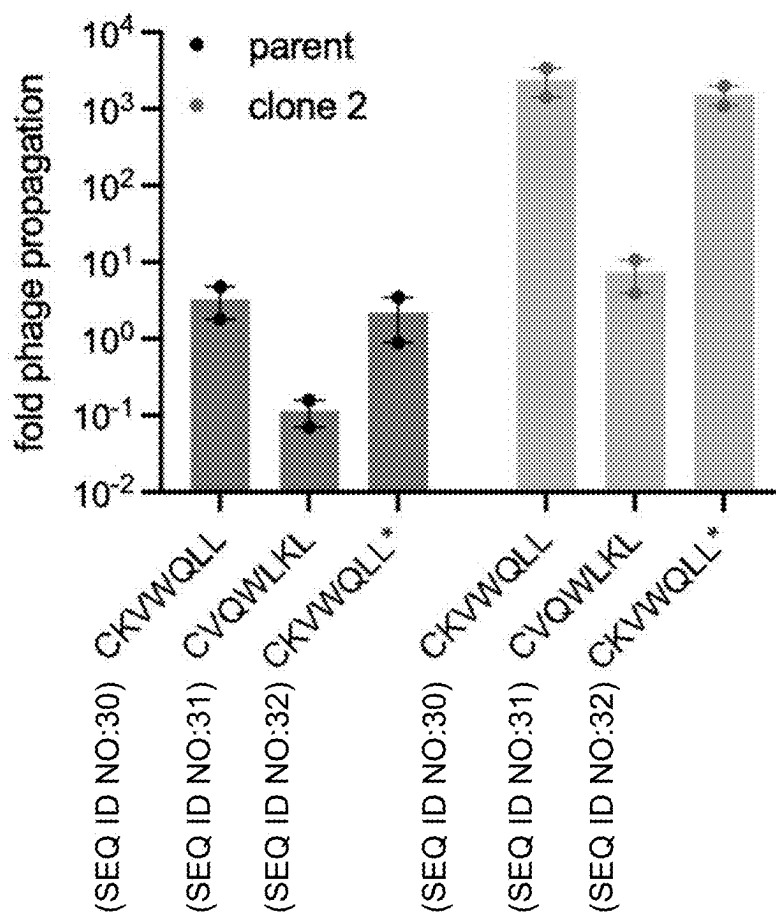


FIG. 8D

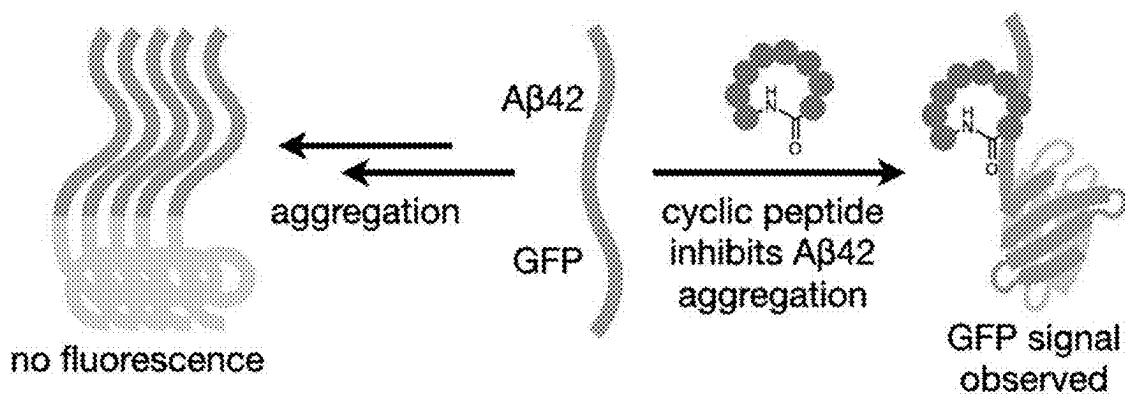


FIG. 8E

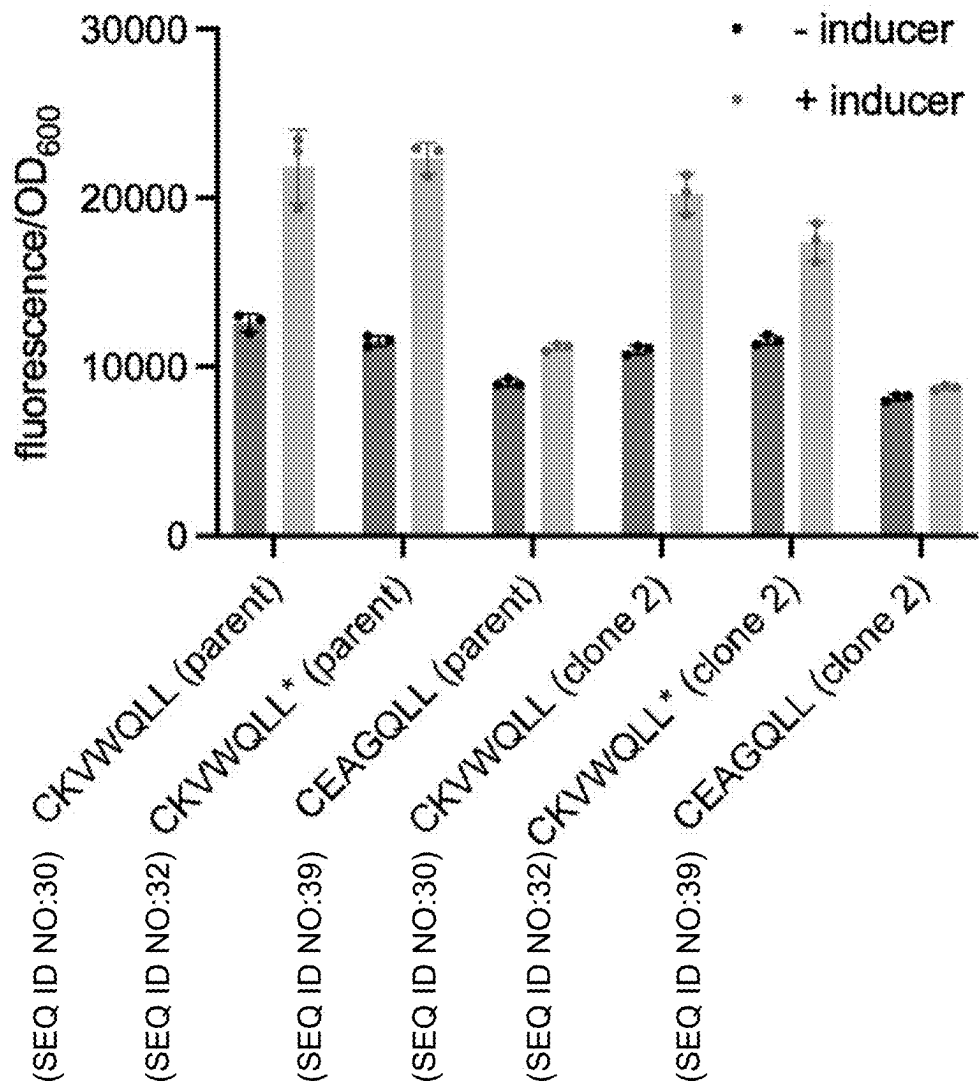


FIG. 8F

top 50 enriched sequences



FIG. 9A

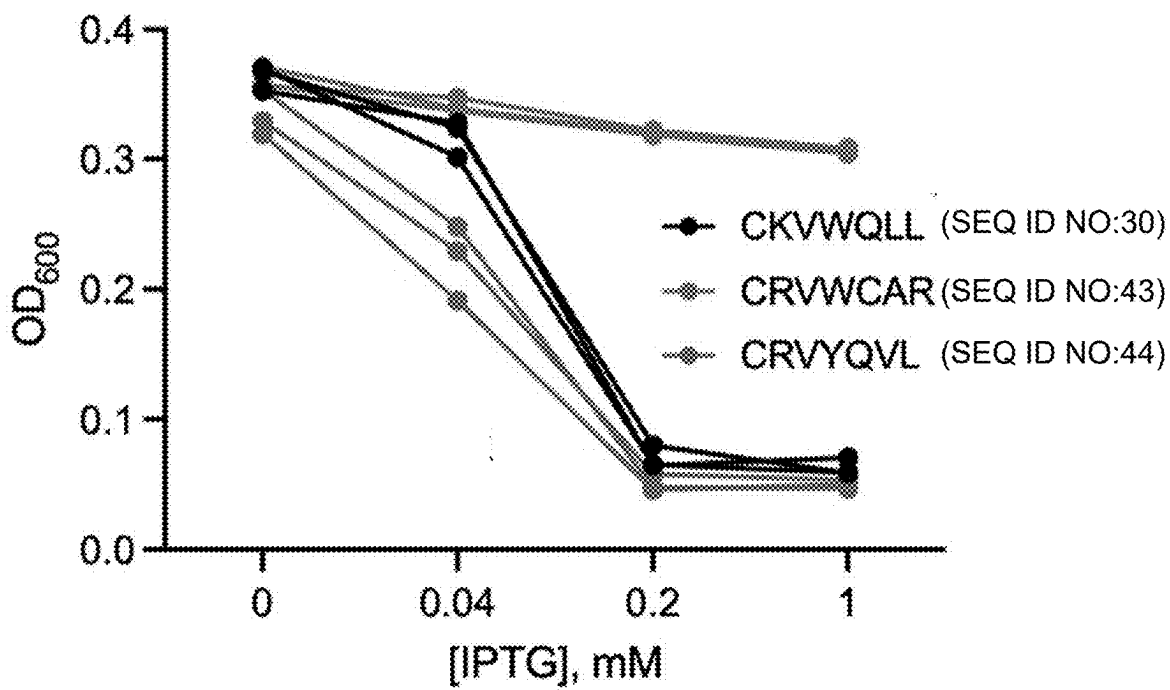
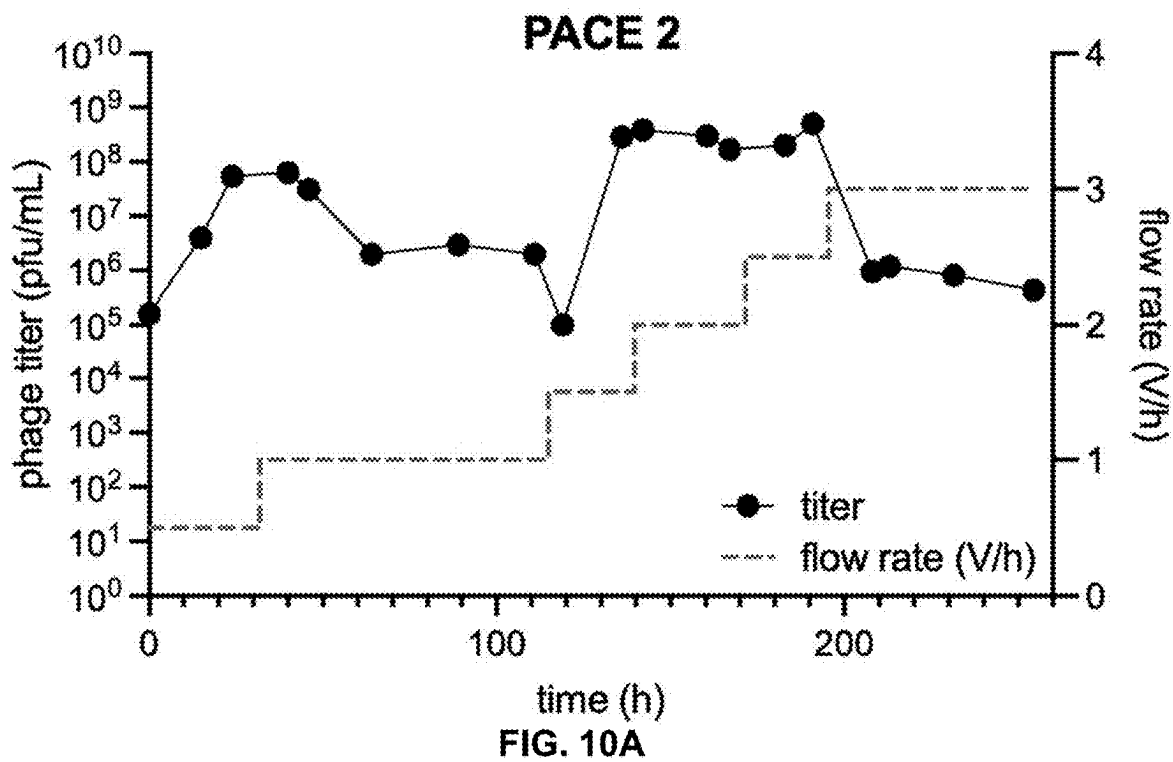


FIG. 9B



	SspC	peptide	SspN							
AA position	10		17	36	38	49	72	80	82	98
parent	L		P	E	R	D	H	Q	L	E
clone 1	F		S	K	K		R	R	V	EKN*
clone 2				K		N		R	V	EKN*
clone 3				K				R	V	EKN*
clone 4				K				R	V	EKN*

FIG. 10B

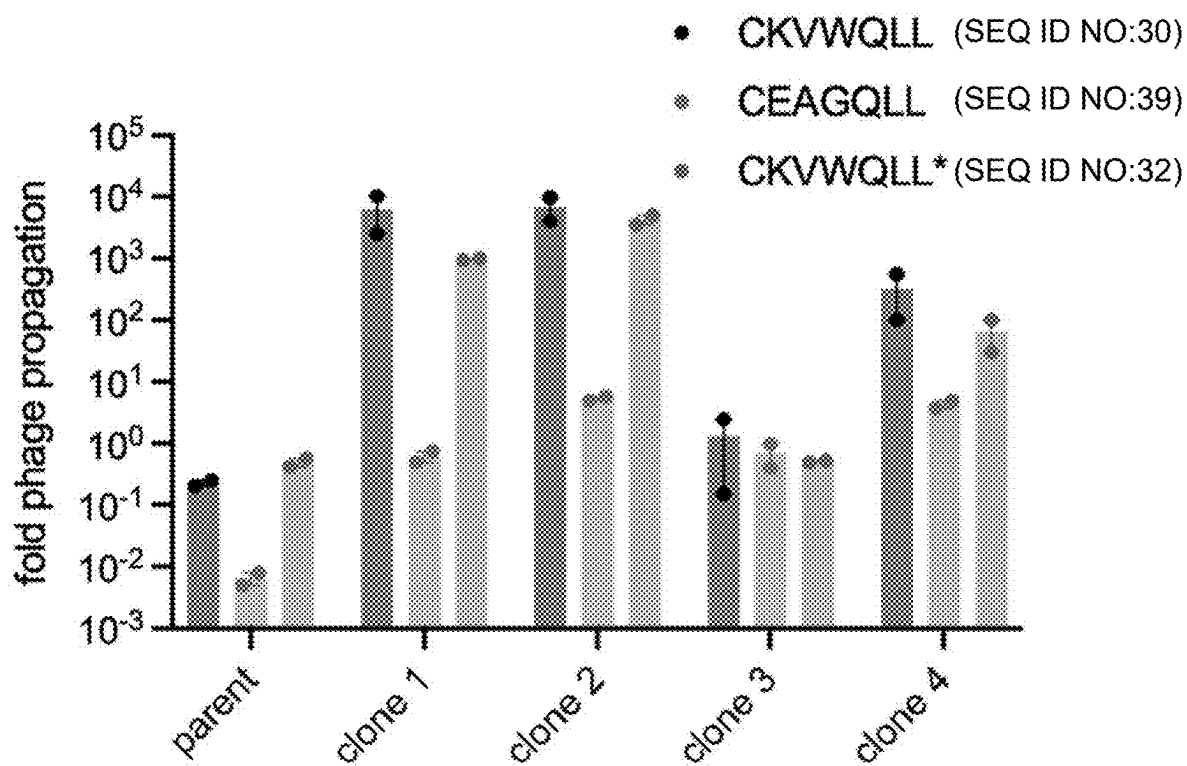


FIG. 10C

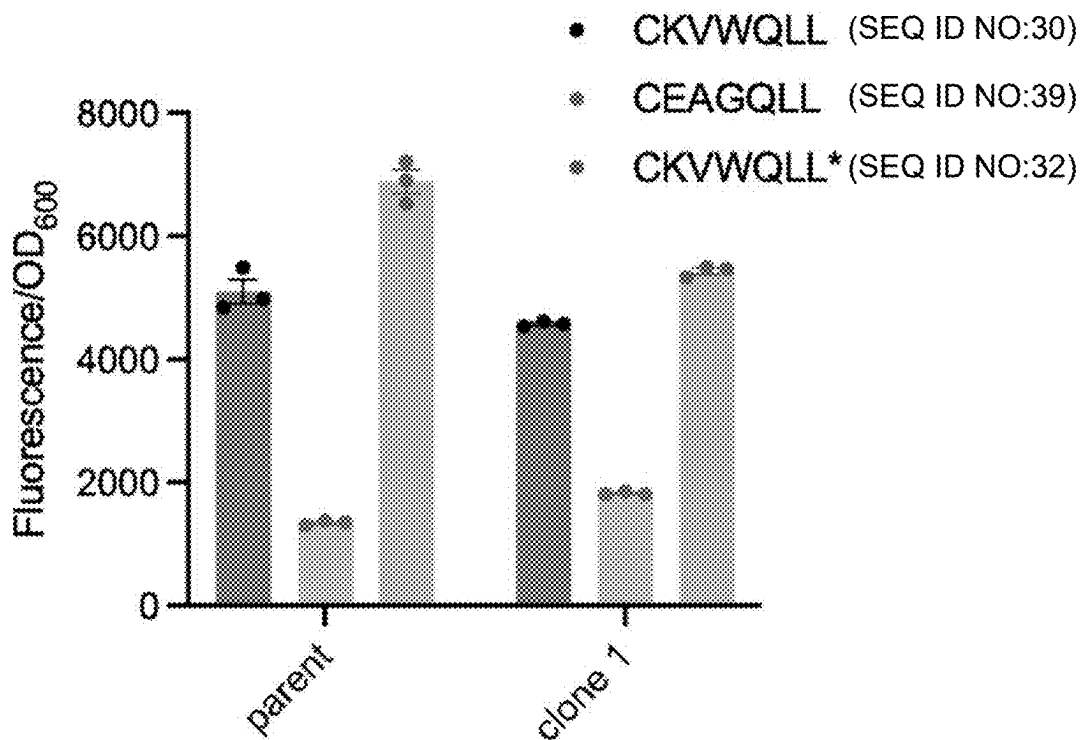


FIG. 10D

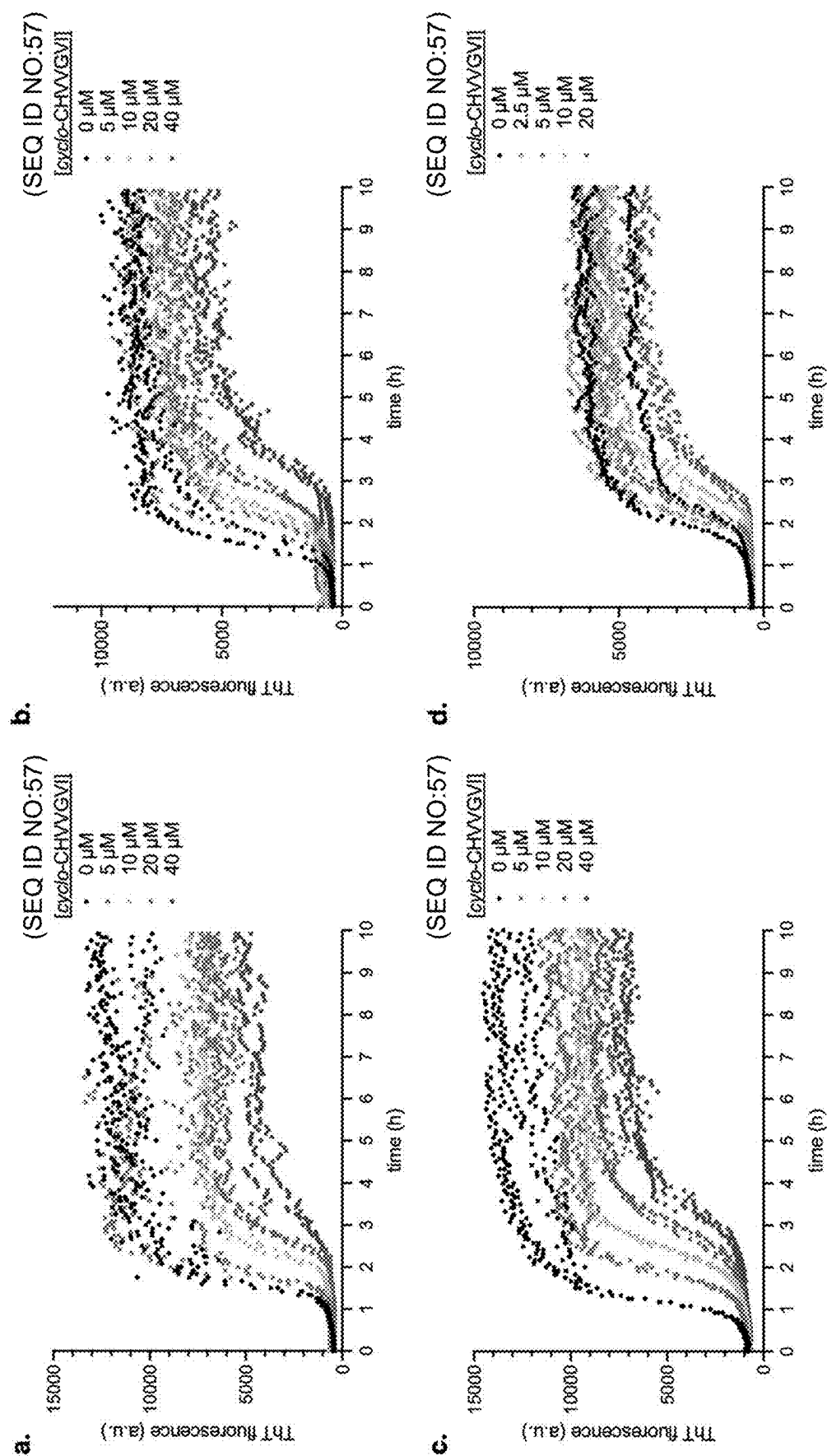


FIG. 11

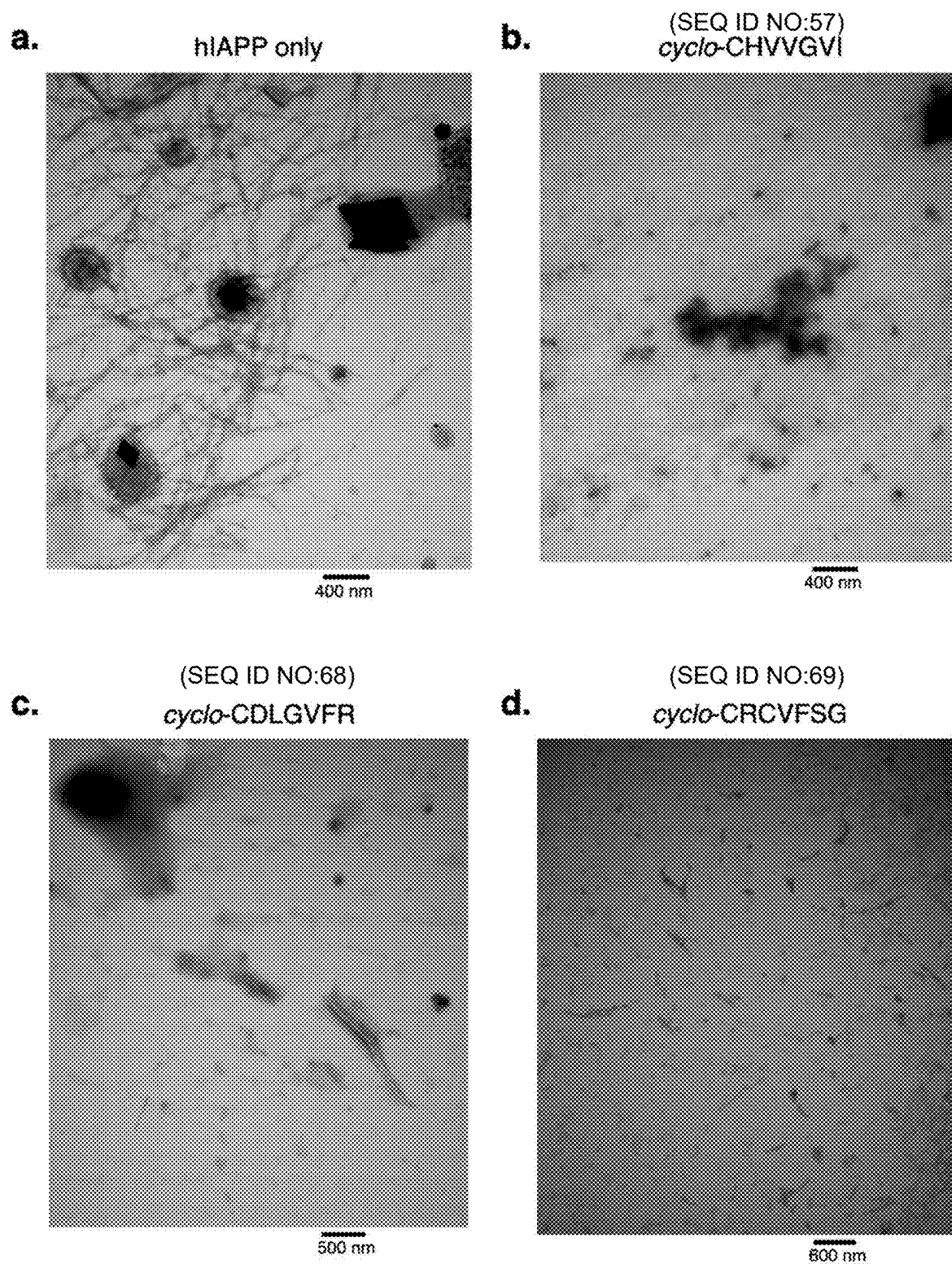


FIG. 12

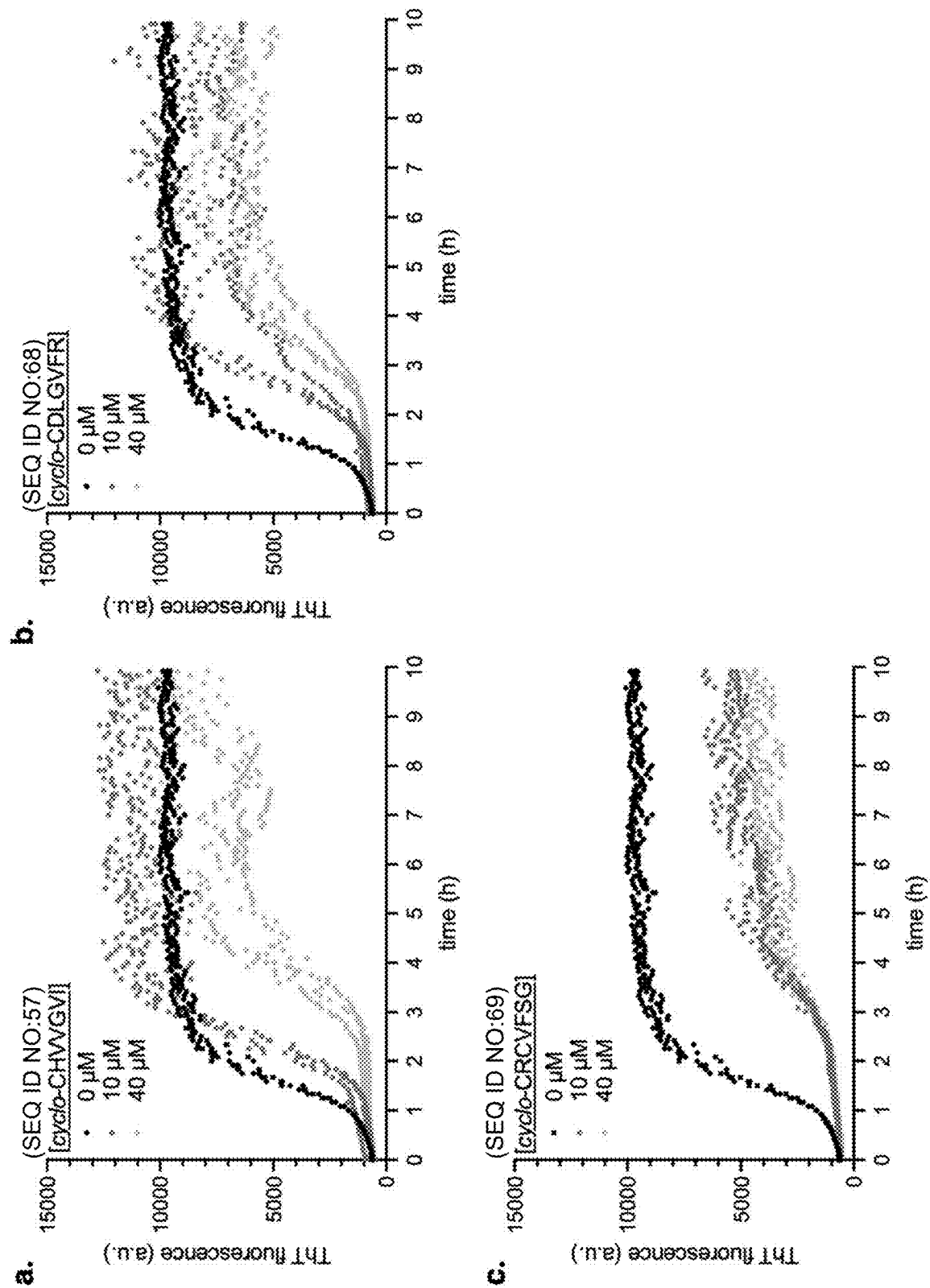


FIG. 13

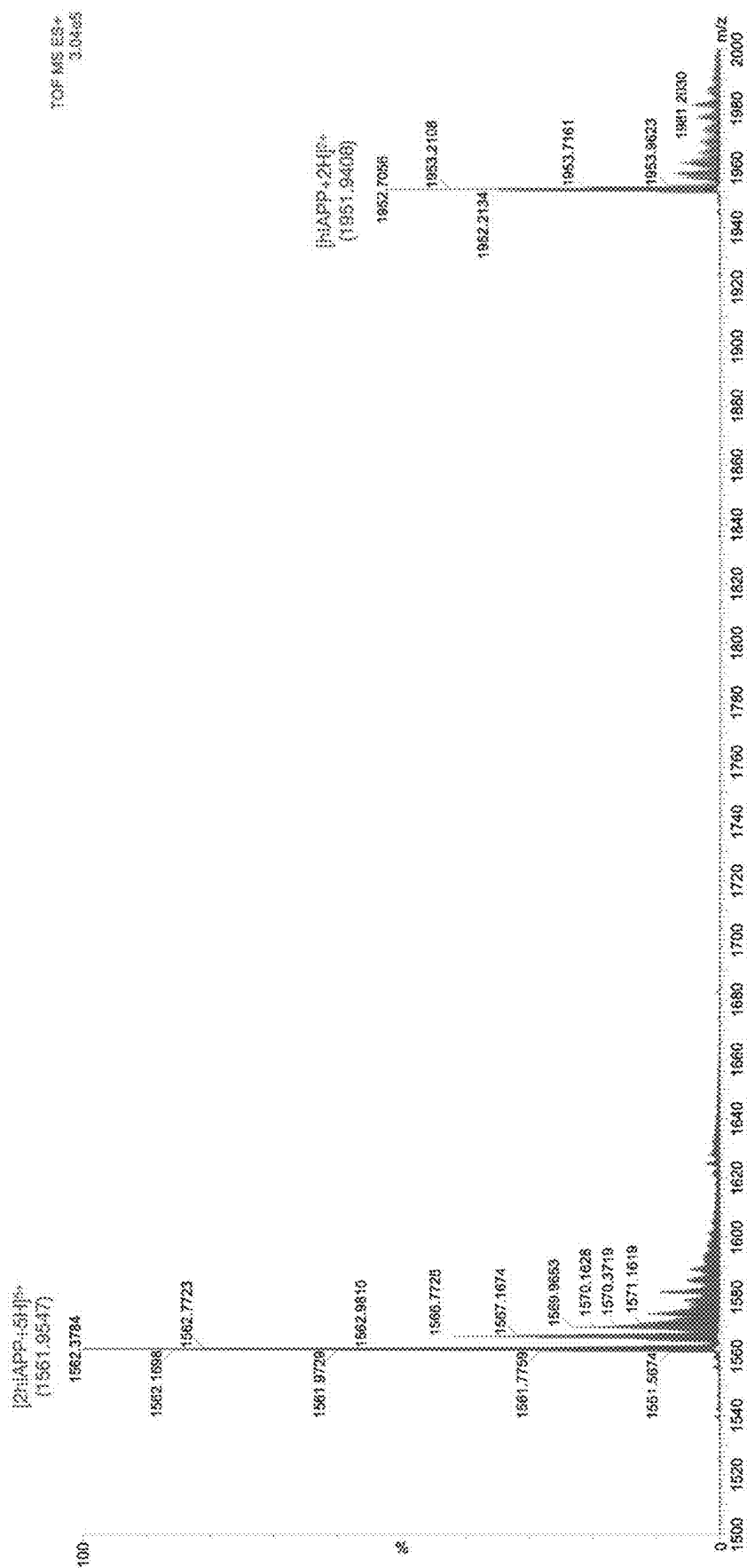


FIG. 14A

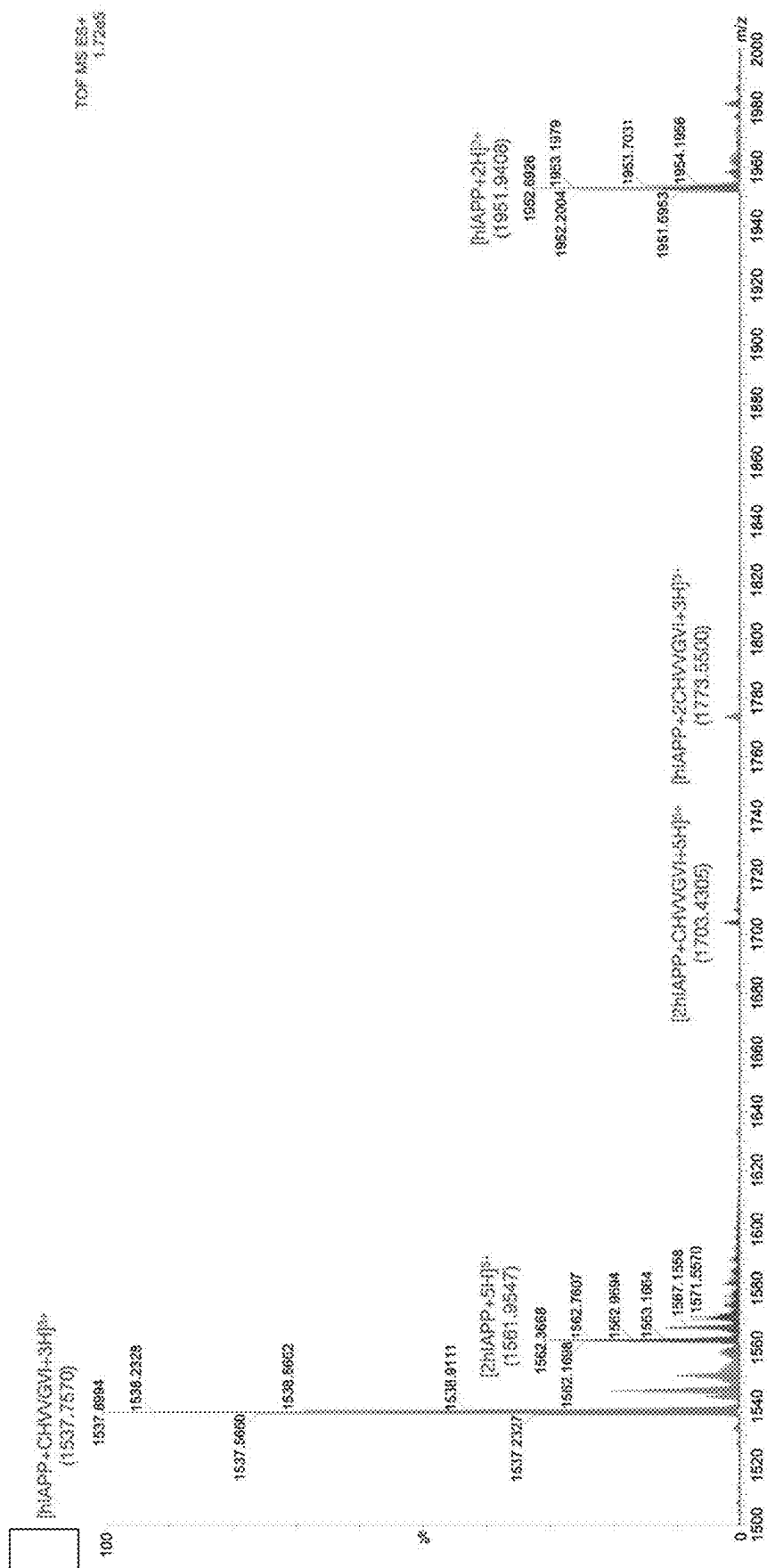


FIG. 14B



FIG. 15A



FIG. 15B

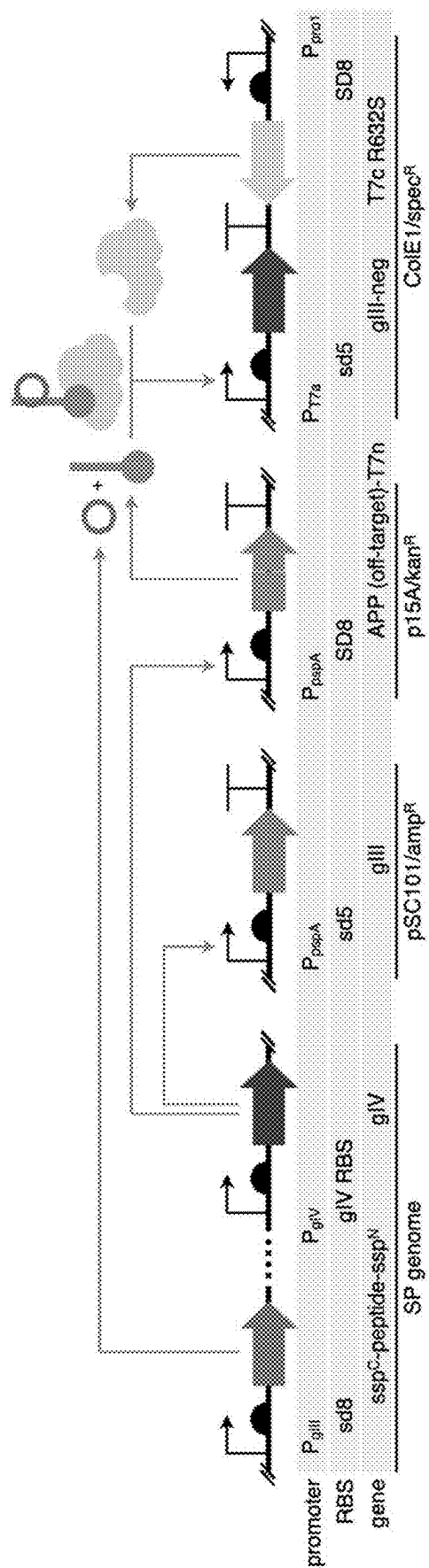


FIG. 16A

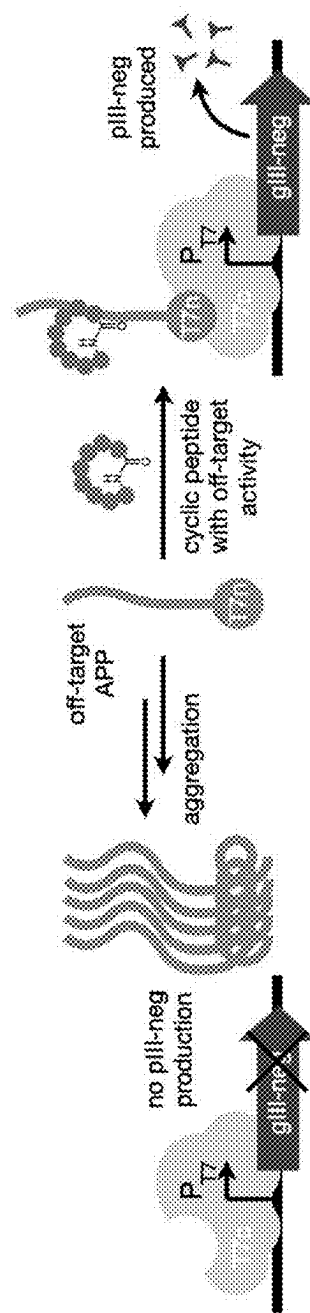


FIG. 16B

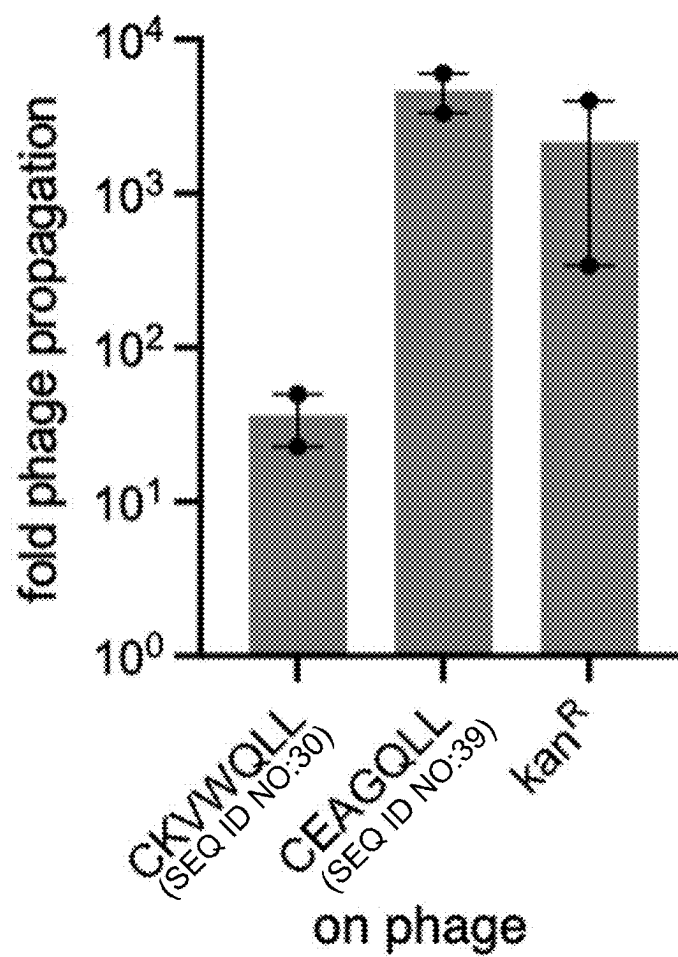


FIG. 16C

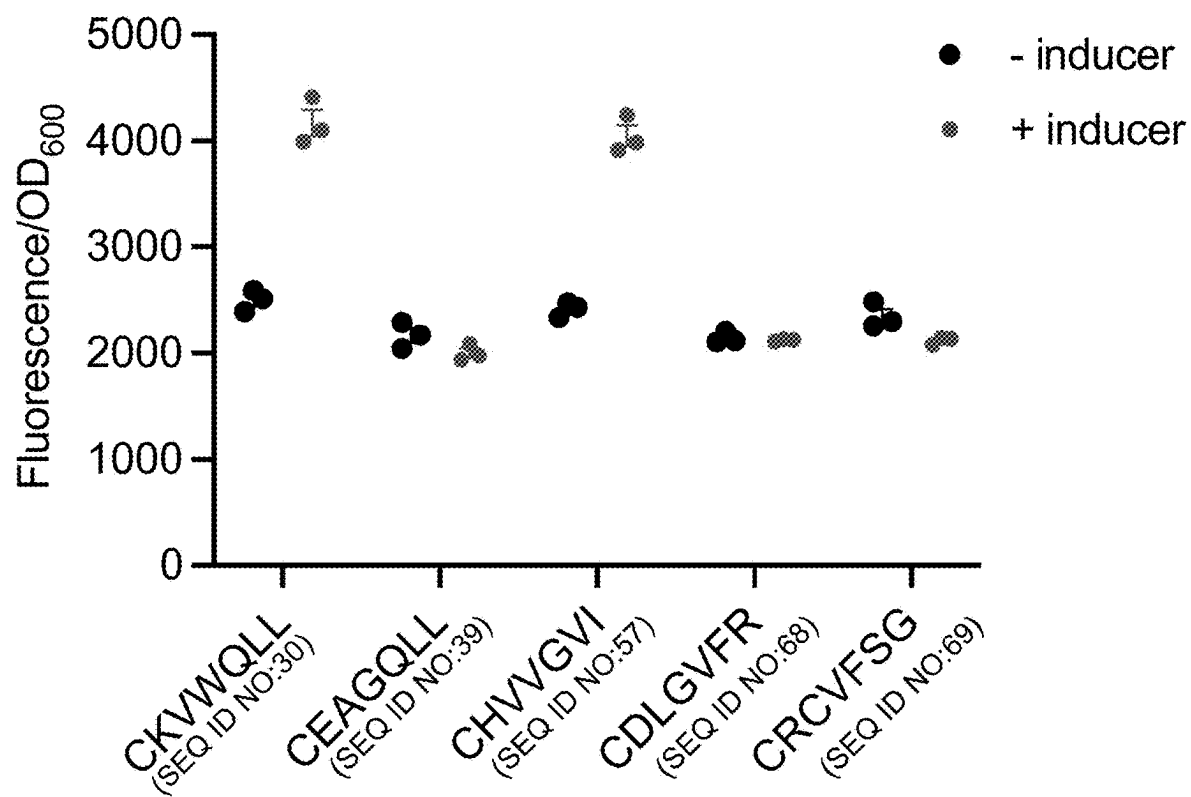


FIG. 17

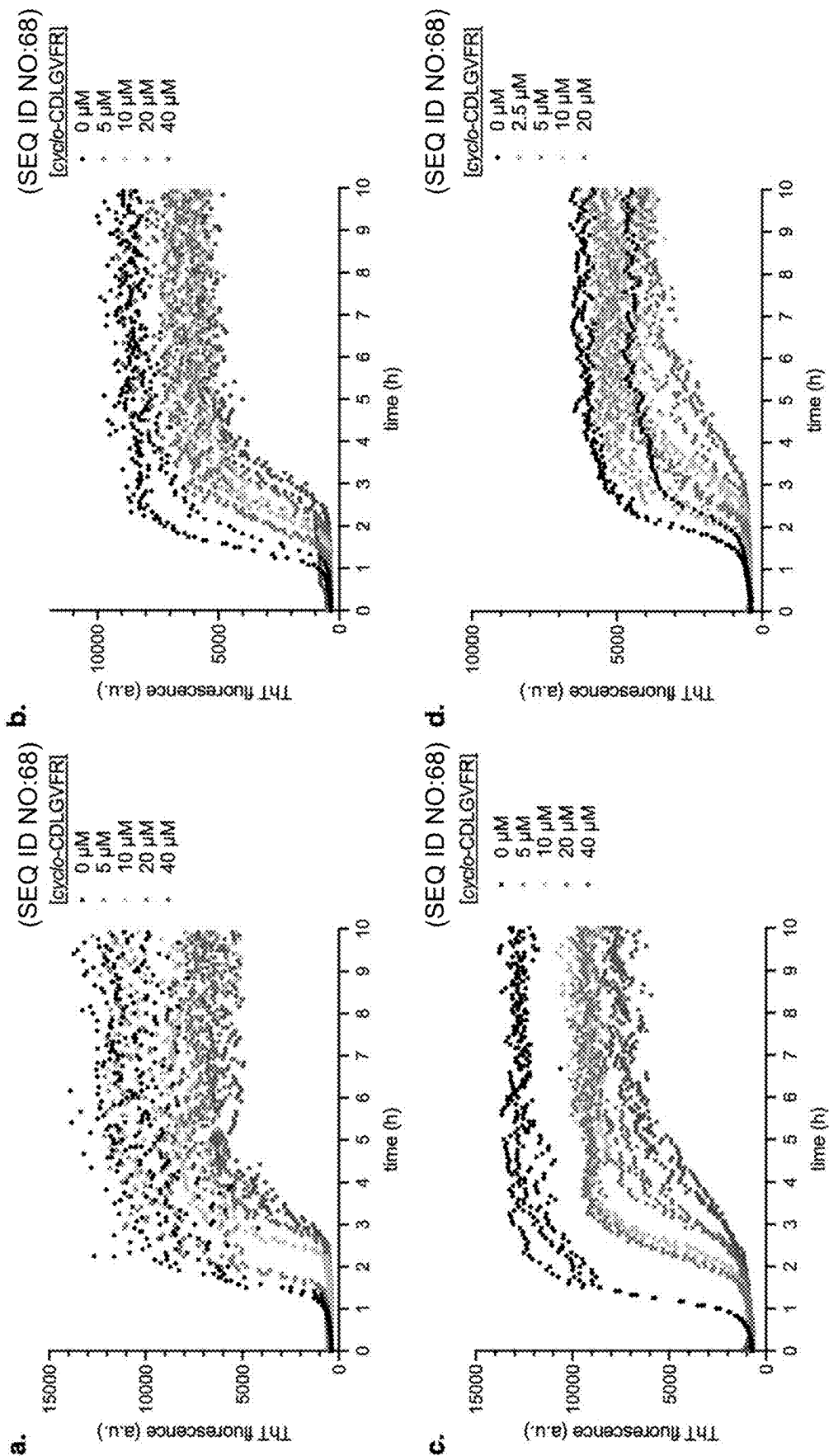


FIG. 18

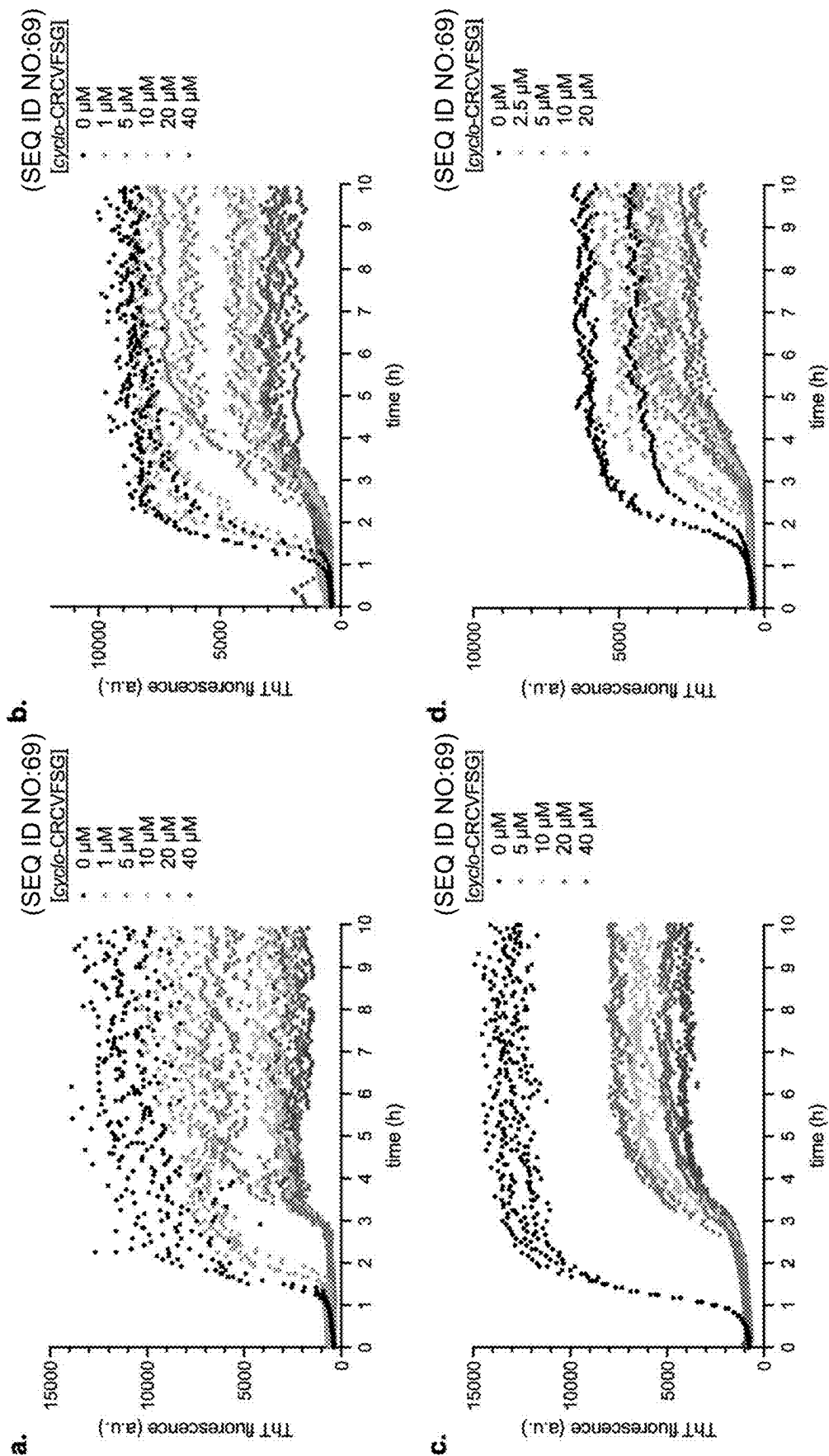


FIG. 19

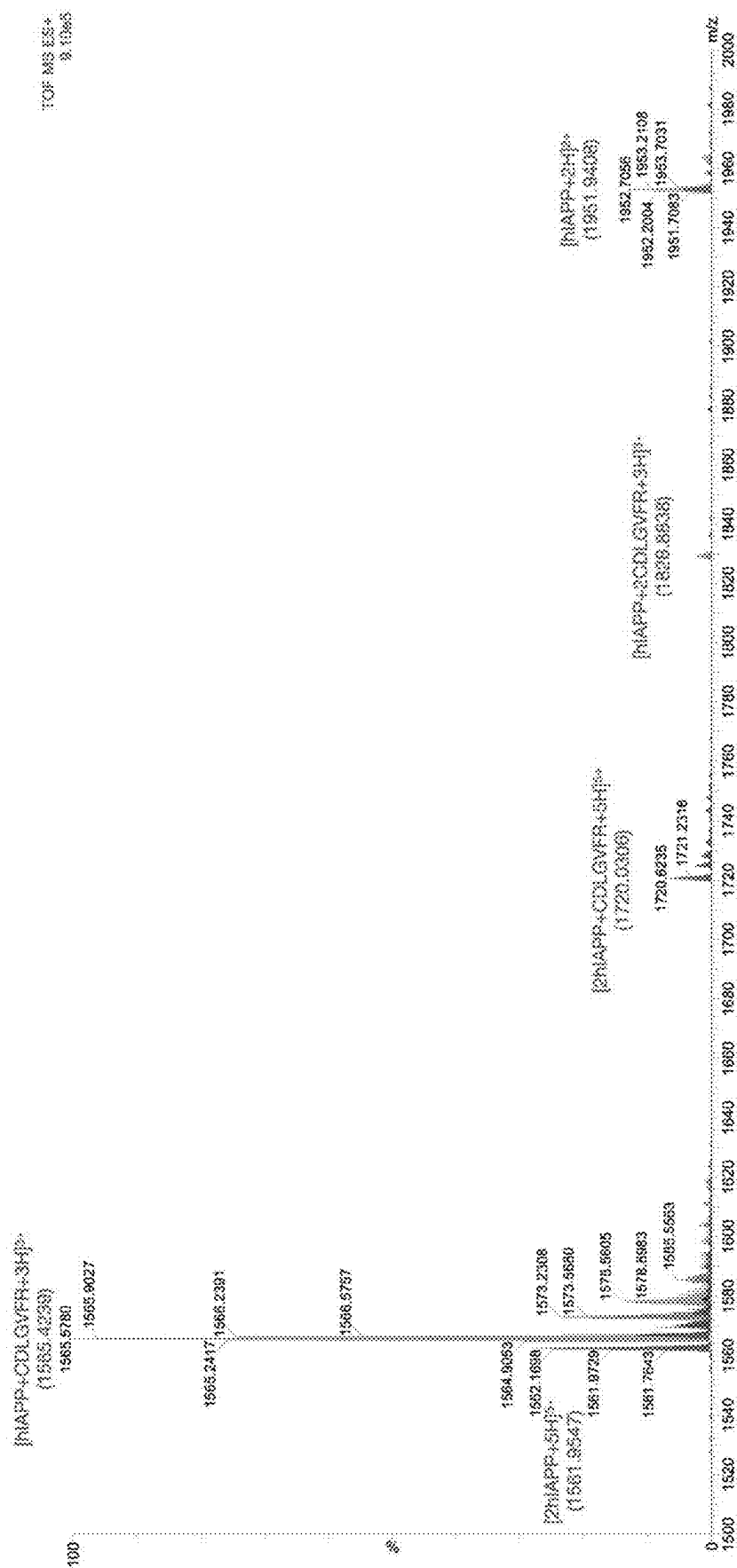


FIG. 20A

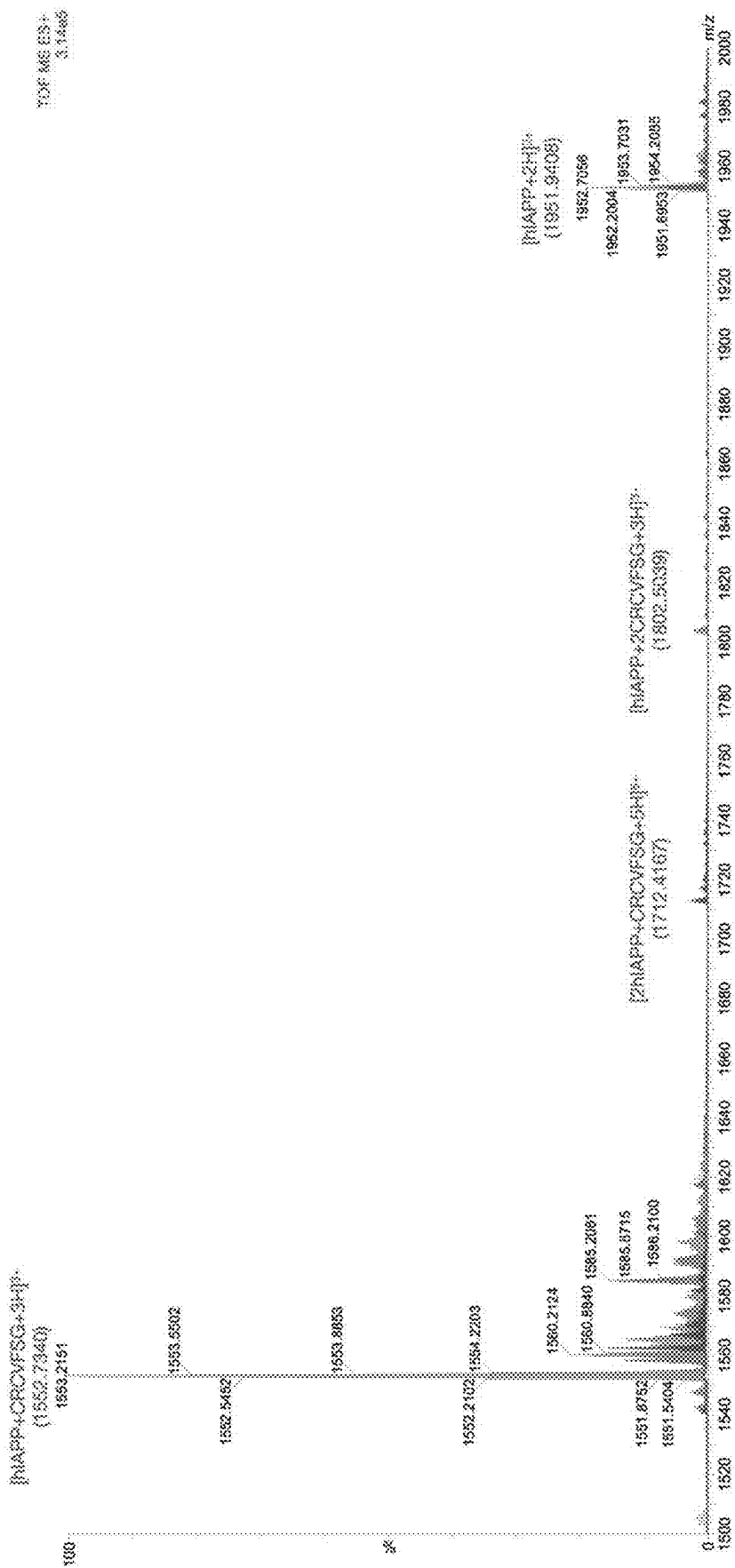


FIG. 20B



FIG. 21A



FIG. 21B

wt. pIII	NKES	424
N-C83	NKES	354
N*-C83	NKES	351
N2-C83	NKES	280
C-domain	NKES	172
C83	NKES	105

FIG. 22

SELECTION SYSTEMS, PEPTIDES DETERMINED THEREWITH, AND METHODS OF USING SAME

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] Priority is hereby claimed to U.S. Provisional Application 63/590,069, filed Oct. 13, 2023, which is incorporated herein by reference in its entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in XML format and is hereby incorporated by reference in its entirety. The XML copy, created on Sep. 27, 2024, is named USPTO—09824557-P230455US02—SEQ_LIST.xml and is 165,377 bytes in size.

FIELD OF THE INVENTION

[0003] The invention is directed to selection systems, such as selection systems for determining peptides that inhibit protein aggregation, peptides determined with the selection systems, and methods of using the selection systems and the peptides.

BACKGROUND

[0004] The aggregation of amyloidogenic proteins has been associated with numerous human diseases.¹ These proteins proceed through a series of misfolded intermediates to ultimately form the ordered, fibrillar structures known as amyloids, motivating the search for compounds that can disrupt their formation. However, structure-guided design of protein misfolding and aggregation inhibitors is complicated by the intrinsically disordered nature of many amyloidogenic species and the dearth of structural information on amyloid precursors.²³

[0005] Inhibitor discovery by high-throughput screening or selection often requires minimal structural information, providing a powerful alternative to rational design.⁴ A recent and exciting example is the screening of small molecule libraries for their effects on the kinetics of amyloid formation in vitro using purified protein.⁵ Such screens have identified promising leads,⁶ but nevertheless can be limited by throughput and the technical challenges of producing high-quality target protein to ensure reproducibility.⁵ Additionally, small molecules can be poorly suited for binding tightly to disordered misfolded species⁷ or strongly inhibiting the protein-protein interactions involved in misfolding and aggregation.⁸

[0006] Macrocyclic peptides have increasingly been recognized as a promising source of chemical probes and potential therapeutics.⁹⁻¹¹ These compounds can bind historically challenging targets, such as large or shallow protein surfaces, and enjoy improved stability and binding potency due to backbone rigidification.^{12,13} Platforms that select genetically encoded cyclic peptide libraries can search through an enormous number of sequences for ones that possess desired bioactivity. For example, mRNA and phage display methods have proven highly successful in identifying macrocyclic inhibitors for numerous targets.^{13,14} However, display techniques are typically constrained to identifying binders of immobilized proteins.¹³ A complementary approach, enabled by the rapidly emerging field of synthetic

biology, involves the genetic or phenotypic selection of ribosomally synthesized cyclic peptide libraries in cells.¹⁵⁻¹⁸ Notably, this strategy can identify cyclic peptides with activities beyond single target binding,¹³ including inhibiting protein aggregation.¹⁹⁻²¹ Inhibitors of α -synuclein aggregation-induced toxicity have been identified by selecting cyclic peptides biosynthesized using the split-intein circular ligation of peptides and proteins (SICLOPPS) method¹⁵ in a yeast synucleinopathy model.¹⁹ Fluorescence-activated cell sorting has been used to select a SICLOPPS library for rescue of a GFP folding reporter in *E. coli* to identify cyclic peptide inhibitors of amyloid- β (A β) and mutant superoxide dismutase aggregation.^{20,21}

[0007] While powerful, these strategies for identifying cyclic peptide protein aggregation inhibitors are coupled to conventional selections such as cellular survival or fluorescence-based sorting, which are time- and labor-intensive and, in the case of cell sorting, require specialized equipment.²² Additionally, selections can suffer from false positives or uncover hits with undesirable properties,¹⁴ such as promiscuity or off-target activity. For example, the majority of hits selected in the Lindquist study were deemed to be false positives either arising from spontaneous mutations in the yeast strain or from off-target activity of the cyclic peptides themselves.¹⁹ Weeding out these false positive and off-target sequences then requires additional experiments, adding to time and labor.

[0008] Selection systems and methods capable of determining peptides, such as cyclic peptides, that inhibit aggregation of proteins are needed, as are peptides that inhibit aggregation.

SUMMARY OF THE INVENTION

[0009] One aspect of the invention is directed to selection systems. In some versions, the selection systems comprise a host cell and a phage library comprising library phages configured to infect the host cell. In some versions, each library phage comprises a library gene configured to express a library peptide in the host cell. In some versions, at least a subset of the library peptides comprise different peptide sequences. In some versions, each library phage is deficient in a replication gene that encodes a replication protein. In some versions, the replication protein is a protein essential for production of infectious progeny phages from the host cell when infected with the library phages. In some versions, the selection systems further comprise a target fusion gene configured to express a target fusion protein in the host cell. In some versions, the target fusion protein comprises a target protein fused to a first subunit of a multi-subunit, promoter-specific RNA polymerase. In some versions, the selection systems further comprise one or more RNA-polymerase subunit genes configured to express one or more additional subunits of the RNA polymerase. In some versions, the selection systems further comprise a selection gene comprising a cognate promoter of the RNA polymerase operationally connected to a coding sequence of a selection protein. In some versions, the selection protein is the replication protein. In some versions, the selection protein is a dominant-negative form of the replication protein.

[0010] In some versions, the RNA polymerase and the cognate promoter are selected from the group consisting of a T7 RNA polymerase and a T7 promoter, a T3 RNA polymerase and a T3 promoter, and an SP6 RNA polymerase and an SP6 promoter, respectively.

[0011] In some versions, the library peptides comprise cyclic peptides. In some versions, at least one of the library peptides binds, is predicted to bind, or is suspected of binding to the target protein. In some versions, at least one of the library peptides binds to the target protein.

[0012] In some versions, the target protein is an aggregation-prone protein. In some versions, the target protein is an amyloidogenic protein.

[0013] In some versions, the replication gene with which the library phages are deficient and the selection protein encoded by the selection gene are selected from the group consisting of: gIII and either pIII or a dominant-negative form of pIII; gIV and either pIV or a dominant-negative form of pIV; gVI and either pVI or a dominant-negative form of pVI; gVII and either pVII or a dominant-negative form of pVII; gVIII and either pVIII or a dominant-negative form of pVIII; and gIX and either pIX or a dominant-negative form of pIX, respectively.

[0014] In some versions, the selection protein is the replication protein. In some versions, the selection protein is the dominant-negative form of the replication protein.

[0015] Another aspect of the invention is directed to methods of selection with a selection system of the invention. The methods can comprise a first selection. In some versions, the first selection comprises contacting a first population of host cells comprising multiple copies of the host cell with the phage library under conditions effective for the library phages to infect the host cells and thereby generate a first population of infected host cells, incubating the first population of infected host cells under conditions effective for production of first progeny phages to thereby produce a first selected phage library comprising the first progeny phages, and harvesting the first selected phage library.

[0016] The methods can further comprise a second selection. In some versions, the second selection comprises harvesting contacting a second population of host cells comprising multiple copies of the host cell with the first selected phage library under conditions effective for the first progeny phages to infect the host cells and thereby generate a second population of infected host cells, incubating the second population of infected host cells under conditions effective for production of second progeny phages to thereby produce a second selected phage library comprising the second progeny phages, and harvesting the second selected phage library.

[0017] In some versions, the first selection, the second selection, or both the first selection and the second selection are performed with a replication protein as a selection protein. In some versions, the first selection, the second selection, or both the first selection and the second selection are performed with a dominant-negative form of a replication protein as the selection protein. In some versions, the first selection is performed with a replication protein as a selection protein, and the second selection is performed with a dominant-negative form of a replication protein as the selection protein.

[0018] Another aspect of the invention is directed to cyclic peptides. In some versions, the cyclic peptides are head-to-tail cyclic peptides. In some versions, the cyclic peptides comprise an amino acid sequence selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HWGVIX_n (SEQ ID NO:3), HVHSYLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n

(SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHHX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLWDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14), NRLDVX_n (SEQ ID NO:15), GLGHGNX_n (SEQ ID NO:16), RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), KVVQLAX_n (SEQ ID NO:19), RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVVGGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26), QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVTAPX_n (SEQ ID NO:29) wherein X is any amino acid and n is an integer from 0-30.

[0019] Another aspect of the invention is directed to methods of reducing aggregation of an aggregation-prone protein. The methods can comprise contacting the aggregation-prone protein with a cyclic peptide of the invention. In some versions, the contacting is performed in vivo. In some versions, the contacting is performed in vitro.

[0020] In some versions, the aggregation-prone protein comprises human islet amyloid polypeptide. In some versions, the cyclic peptide has a sequence selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HVVGVI_n (SEQ ID NO:3), HVHSYLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n (SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHHX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLWDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14), NRLDVX_n (SEQ ID NO:15), and GLGHGNX_n (SEQ ID NO:16) wherein X is any amino acid and n is an integer from 0-30. In some versions the contacting is performed in a subject with type 2 diabetes.

[0021] In some versions, the aggregation-prone protein comprises amyloid-β42. In some versions, the cyclic peptide has a sequence selected from the group consisting of RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), KVVQLAX_n (SEQ ID NO:19), RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVVGGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26), QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVTAPX_n (SEQ ID NO:29) wherein X is any amino acid and n is an integer from 0-30. In some versions, the contacting is performed in a subject with Alzheimer's disease.

[0022] The objects and advantages of the invention will appear more fully from the following detailed description of the preferred embodiment of the invention made in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0024] FIG. 1. Strategy for cyclic peptide discovery. Selection phage (SP) carrying a library of genetically encoded cyclic peptides are used to infect host *E. coli* cells transformed with an "accessory plasmid." The accessory

plasmid encodes a selection that links a user defined inhibitory activity to expression of phage protein pill, which is required for phage reproduction. Upon infection, the cyclic peptide library member is expressed from the transduced SP genome and, if active, triggers production of pIII from the accessory plasmid. As a result, only SP that carry active library members propagate, allowing rapid enrichment of desired cyclic peptide sequences.

[0025] FIGS. 2A-2E. Design of a selection for cyclic peptide protein aggregation inhibitors. (FIG. 2A) Cartoon overview of a split T7 RNAP folding reporter strategy for identifying protein aggregation inhibitors. APP=aggregation-prone protein, T7n and T7c=N- and C-terminal halves of split-T7 RNAP, respectively. (FIG. 2B) Diagram of the genetic circuits that constitute the protein aggregation inhibitor selection. (FIG. 2C) Propagation activity of kanR-encoding phage on host cells encoding the APP-T7n selection, where the APP is either A β 42 or its non-aggregating F19S/L34P mutant. (FIG. 2D) Comparison of phage propagation activities of SP carrying A β 42 aggregation inhibitor cyclo-CKVWQLL (SEQ ID NO:30) or scrambled negative control cyclo-CVQWLKL (SEQ ID NO:31) on host cells encoding the A β 42-T7n selection. (FIG. 2E) Effect of the mutation of key residues in cyclo-CKVWQLL (SEQ ID NO:30) on SP propagation on host cells encoding the A β 42-T7n selection. (FIGS. 2C-2E) Data reflects mean and standard error (s.e.) of two (FIGS. 2C and 2D) or three (FIG. 2E) biological replicates. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage (10⁹) used to infect the culture. calculated as the number of phage generated from an infected culture divided by the number of phage (10⁹) used to infect the culture.

[0026] FIGS. 3A-3F. Selection of cyclic peptide A β 42 aggregation inhibitors. (FIG. 3A) Graphical overview of selection procedure. (FIG. 3B) Phage titers (solid lines, plotted on left y-axis) and propagation activity (dashed lines, plotted on right y-axis) of successive rounds of selection of an SP-encoded cyclo-CX₃QLX library on selection cells encoding A β 42 as the target APP. “In” refers to the input SP pool used to infect selection cells and “out” refers to the output SP pool obtained after selection, for a given round of selection. pfu=plaque-forming units. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage used to infect the culture. (FIG. 3C) Sanger sequencing of cyclic peptides sequences encoded by clonal phage isolated from round 7 of the selection. Residues in gray italics were fixed in the starting library. (FIG. 3D) Phage titers (solid lines, plotted on left y-axis) and propagation activity (dashed lines, plotted on right y-axis) of successive rounds of selection of an SP-encoded cycloCX6 library on selection cells encoding A β 42 as the target APP. “In” refers to the input SP pool used to infect selection cells and “out” refers to the output SP pool obtained after selection, for a given round of selection. pfu=plaque-forming units. (FIG. 3E) Top 10 enriched cyclic peptide sequences after 7 rounds of selection on A β 42 as determined by high-throughput sequencing analysis. (FIG. 3F) Propagation activity of clonal phage encoding enriched sequences CRVWCAR (SEQ ID NO:43) and CRVYQVL (SEQ ID NO:44) on host cells encoding the A β 42-T7n selection. Phage encoding active sequence CKVWQLL (SEQ ID NO:30) is shown for comparison. Data reflect

mean and s.e. of two biological replicates. (FIG. 3G) Propagation activity of clonal phage encoding single alanine substitutions of the CRVWCAR (SEQ ID NO:43) sequence on host cells encoding the A β 42-T7n selection. Data reflect mean and s.e. of three biological replicates. (FIG. 3H) Evaluation of cyclic peptide sequences enriched by selection on A β 42-T7n expressed from plasmid using the A β 42-GFP folding reporter. CKVWQLL (SEQ ID NO:30) is shown for comparison. Cyclic peptides are induced at 0.04 mM IPTG. Data reflects mean and s.e. of three biological replicates. (FIGS. 3D, 3F, and 3G) Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage used to infect the culture. For (FIGS. 3F and 3G), 105 input phage were used to infect.

[0027] FIGS. 4A-4J. Selection of cyclic peptide inhibitors of hIAPP aggregation. (FIG. 4A) Phage titer (solid line, plotted on left y-axis) and propagation activity (dashed line, plotted on right y-axis) of successive rounds of selection of an SP-encoded SICLOPPS library on selection cells targeting hIAPP aggregation. “In” refers to the input SP pool used to infect selection cells and “out” refers to the output SP pool obtained after selection, for a given round of selection. pfu=plaqueforming units. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage used to infect the culture. (FIG. 4B) Propagation activity of SP pools from each round of the selection shown in (FIG. 4A) where input phage titers are normalized to 105 pfu/mL. Data reflect two biological replicates plotted as individual values. (FIG. 4C) Top 10 enriched cyclic peptide sequences after 5 rounds of selection on hIAPP as determined by high-throughput sequencing analysis. (FIG. 4D) Structure of cyclo-CHVVGVI (SEQ ID NO:57). (FIG. 4E) Effect of varying concentrations of chemically synthesized cyclo-CHVVGVI (SEQ ID NO:57) (left) or linear CHVVGVI (SEQ ID NO:67) (right) on aggregation of 10 μ M hIAPP measured by ThT fluorescence. Assay was run in phosphate-buffered saline (PBS) pH 6.9 at 32° C. under quiescent conditions. Data reflect plots of individual values of three technical replicates. The hIAPP alone (0 μ M peptide) data is the same in both plots. (FIG. 4F) Effect of cyclo-CHVVGVI (SEQ ID NO:57) on lag time (tag) of hIAPP aggregation. Data reflects tag fits from four biological replicates (4 separate ThT assays; see FIG. 11) plotted as individual values. P-values (two-tailed Student’s t-test between 0 μ M peptide and each peptide concentration): *, P $\leq$ 0.05; **, P $\leq$ 0.01. (FIGS. 4G and 4H) TEM analysis of 10 μ M hIAPP incubated in the presence (FIG. 4G) of 40 μ M cyclo-CHVVGVI (SEQ ID NO:57) or alone (FIG. 4H) for for 2 h. (i-j) TEM analysis of 10 μ M hIAPP in the presence (FIG. 4I) of 40 μ M cyclo-CHVVGVI (SEQ ID NO:57) or alone (FIG. 4J) for 14 h. (FIGS. 4G-4J) For compatibility with TEM, hIAPP incubations were run in 25 mM sodium phosphate buffer, pH 7.4; see FIG. 13 for hIAPP aggregation kinetics under TEM assay conditions. Scale bars=200 nm.

[0028] FIGS. 5A-5E. Negative selection to remove promiscuous cyclic peptide sequences. (FIG. 5A) Propagation activity of the SP pool from round 5 of positive selection for hIAPP inhibition on selection cells encoding hIAPP, A β 42, or an insoluble antibody fragment (scFv) fused to T7n. (FIG. 5B) Phage titer (solid line, plotted on the left y-axis) and propagation activity (dashed line, plotted on the right y-axis) of successive rounds of selection of the hIAPP round 5 SP pool on A β 42-T7n negative selection cells. “In” refers to the

input SP pool used to infect selection cells and “out” refers to the output SP pool obtained after selection, for a given round of selection. pfu=plaque-forming units. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage used to infect the culture. (FIG. 5C) Propagation activity of pre- and post-negative selection SP pools on selection cells encoding hIAPP or A β 42 as the target APP. (FIG. 5D) Top 10 enriched cyclic peptide sequences after 4 rounds of negative selection as determined by high-throughput sequencing analysis. (FIG. 5E) Propagation of clonal phage encoding cyclic peptide sequences enriched by either positive selection on hIAPP only (CHVVGVI (SEQ ID NO:57)) or by hIAPP positive selection followed by negative selection against A β 42 inhibitors (CDLGVFR (SEQ ID NO:68) and CRCVFSG (SEQ ID NO:69)) on selection cells encoding hIAPP or A β 42 as the target APP. (FIGS. 5A, 5C, and 5E) Data reflects the mean and s.e. of two biological replicates. Propagations used an input of 105 phage.

[0029] FIGS. 6A-6J. Cyclic peptides identified by negative selection inhibit hIAPP aggregation. (FIGS. 6A and 6B) Structures of cyclo-CDLGVFR (SEQ ID NO:68) and cyclo-CRCVFSG (SEQ ID NO:69). (FIG. 6C) Effect of varying concentrations of chemically synthesized cyclo-CDLGVFR (SEQ ID NO:68) (left) or linear CDLGVFR (SEQ ID NO:78) (right) on aggregation of 10 μ M hIAPP measured by ThT fluorescence. Assay was run in phosphate-buffered saline (PBS) pH 6.9 at 32° C. under quiescent conditions. Data reflect plots of individual values of three technical replicates. The hIAPP alone (0 μ M peptide) data is the same in both plots. (FIG. 6D) Effect of cyclo-CDLGVFR (SEQ ID NO:68) on lag time (t_{lag}) of hIAPP aggregation. Data reflects t_{lag} fits from four biological replicates (4 separate ThT assays; see FIG. 18) plotted as individual values. P values (two-tailed Student's t-test between 0 μ M peptide and each peptide concentration): **, P $\leq$ 0.01. (FIG. 6E) Effect of varying concentrations of chemically synthesized cyclo-CRCVFSG (SEQ ID NO:69) (left) or linear CRCVFSG (SEQ ID NO:79) (right) on aggregation of 10 μ M hIAPP measured by ThT fluorescence. Assay was run in phosphate-buffered saline (PBS) pH 6.9 at 32° C. under quiescent conditions. Data reflect plots of individual values of three technical replicates. The hIAPP alone (0 μ M peptide) data is the same as in (FIG. 6C). (FIG. 6F) Effect of cyclo-CRCVFSG (SEQ ID NO:69) on lag time (t_{lag}) of hIAPP aggregation. Data reflects t_{lag} fits from four biological replicates (4 separate ThT assays; see FIG. 19) plotted as individual values. P-values (two-tailed Student's t-test between 0 μ M peptide and each peptide concentration): **, P $\leq$ 0.01. (FIGS. 6G and 6H) TEM analysis of 10 μ M hIAPP incubated in the presence of 40 μ M cyclo-CDLGVFR (SEQ ID NO:68) (FIG. 6G) or cyclo-CRCVFSG (SEQ ID NO:69) (FIG. 6H) for 2 h. (i-j) TEM analysis of 10 μ M hIAPP incubated in the presence of 40 μ M cyclo-CDLGVFR (SEQ ID NO:68) (FIG. 6G) or cyclo-CRCVFSG (SEQ ID NO:69) (FIG. 6H) for 14 h. (FIGS. 6G-6J) For compatibility with TEM, hIAPP incubations were run in 25 mM sodium phosphate buffer, pH 7.4; see FIG. 13 for hIAPP aggregation kinetics under TEM assay conditions. Scale bars=200 nm.

[0030] FIG. 7. Mechanism of head-to-tail cyclic peptide formation by SICLOPPS.

[0031] FIGS. 8A-8F. (FIG. 8A) Propagation of parent SP encoding CKVWQLL (SEQ ID NO:30) or scrambled control CVQWLKL (SEQ ID NO:31) on host cells carrying the

A β 42 selection. Data reflects mean and standard error (s.e.) of 2 biological replicates. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage (105) used to infect the culture. (FIG. 8B) Phage titers from phage-assisted continuous evolution (PACE) of SP encoding the SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO:30) on selection cells encoding A β 42-T7n using AP pTW006aP1a. pfu=plaque-forming units. (FIG. 8C) Mutations from clonal SP isolated from the last timepoint of PACE 1. (FIG. 8D) Evaluation of clonal SP activity on selection cells encoding A β 42-T7n using AP pTW357b. Active (CKVWQLL ((SEQ ID NO:30))), scrambled control (CVQWLKL ((SEQ ID NO:31))), and intein splicing-inactivated (CKVWQLL* (SEQ ID NO:32)) cyclic peptide sequences were evaluated. Data reflects mean and standard error (s.e.) of 2 biological replicates. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage (10⁵) used to infect the culture. (FIG. 8E) Overview of the A β 42-GFP assay for detecting A β 42 aggregation inhibition. (FIG. 8F) Evaluation of either parent or evolved (clone 2) SICLOPPS cyclic peptide precursor expressed from plasmid using the A β 42-GFP folding reporter. Active (CKVWQLL (SEQ ID NO:30)), negative control (CEAGQLL (SEQ ID NO:39)), and intein splicing-inactivated (CKVWQLL* (SEQ ID NO:32)) cyclic peptide sequences were evaluated. Cyclic peptides are induced at 0.04 mM IPTG. Data reflects mean and s.e. of 3 biological replicates.

[0032] FIGS. 9A and 9B. (FIG. 9A) Sequence logos generated from the top 50 enriched sequences from selection of a phage-encoded cyclo-CX₆ library for A β 42-T7n inhibition. (FIG. 9B) OD₆₀₀ of cells expressing SICLOPPS precursors for cyclo-CKVWQLL (SEQ ID NO:30), cyclo-CRVWCAR (SEQ ID NO:43), and cyclo-CRVYQVL (SEQ ID NO:44) upon induction by IPTG. Data reflects 3 biological replicates plotted as individual values.

[0033] FIGS. 10A-10D. (FIG. 10A) Phage titers from PACE of SP encoding the SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO:30) on selection cells encoding A β 42-T7n using AP pTW357d5. pfu=plaque-forming units. (FIG. 10B) Mutations from clonal SP isolated from the last timepoint of PACE 2. (FIG. 10C) Evaluation of clonal SP activity on selection cells encoding A β 42-T7n using AP pTW357d5. Active (CKVWQLL (SEQ ID NO:30)), negative control (CEAGQLL (SEQ ID NO:39)), and intein splicing-inactivated (CKVWQLL* (SEQ ID NO:32)) cyclic peptide sequences were evaluated. Data reflects mean and standard error (s.e.) of two biological replicates. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage (10⁵) used to infect the culture. (FIG. 10D) Evaluation of either parent or evolved (PACE 2 clone 1) SICLOPPS cyclic peptide precursor expressed from plasmid using the A β 42-GFP folding reporter. Active (CKVWQLL (SEQ ID NO:30)), negative control (CEAGQLL (SEQ ID NO:39)), and intein splicing-inactivated (CKVWQLL* (SEQ ID NO:32)) cyclic peptide sequences were evaluated. Cyclic peptides are induced at 0.04 mM IPTG. Data reflects mean and s.e. of 3 biological replicates.

[0034] FIG. 11. Replicates of hIAPP aggregation assays with cyclo-CHVVGVI (SEQ ID NO:57) used to calculate t_{lag} values shown in FIG. 4F. Each graph reflects plots of individual values of three technical replicates.

[0035] FIG. 12. TEM images with expanded fields of view from of 10 μ M hIAPP incubated (a) alone or in the presence of (b) 40 μ M cyclo-CHVVGVI (SEQ ID NO:57), (c) 40 μ M cyclo-CDLGVFR (SEQ ID NO:68), or (d) 40 μ M cyclo-CRCVFSG (SEQ ID NO:69) for 2 h. hIAPP incubations were run in 25 mM sodium phosphate buffer, pH 7.4; see FIG. 13 for hIAPP aggregation kinetics under TEM assay conditions.

[0036] FIG. 13. Effect of (a) cyclo-CHVVGVI (SEQ ID NO:57), (b) cyclo-CDLGVFR (SEQ ID NO:68), or (c) cyclo-CRCVFSG (SEQ ID NO:69) on aggregation of 10 μ M hIAPP measured by ThT fluorescence. Assay was run in 25 mM sodium phosphate buffer, pH 7.4 at 32° C. under quiescent conditions. Each graph reflects plots of individual values of three technical replicates. The hIAPP alone (0 μ M peptide) data is the same in panels (a-c).

[0037] FIGS. 14A and 14B. Native mass spectra of 10 μ M hIAPP alone (FIG. 14A) or in the presence of 40 μ M cyclo-CHVVGVI (SEQ ID NO:57) (FIG. 14B). Peak assignments are shown in blue with the theoretical masses given in parentheses. For the full spectra, see FIGS. 15A and 15B.

[0038] FIGS. 15A and 15B. Full native mass spectra of 20 μ M hIAPP alone (FIG. 15A) or in the presence of 40 μ M cyclo-CHVVGVI (SEQ ID NO:57) (FIG. 15B). Peak assignments are shown in blue with the theoretical masses given in parentheses.

[0039] FIGS. 16A-16C. Negative selection to remove off-target cyclic peptides. (FIG. 16A) Diagram of genetic circuits constituting the negative selection. (FIG. 16B) Cartoon depiction of negative selection strategy using dominant-negative pIII variant pIII-neg. (FIG. 16C) Propagation of SP on negative selection cells that produce pIII-neg upon inhibition of A β 42-T7n aggregation. The tested SP encode cyclo-CKVWQLL (SEQ ID NO:30) (A β 42 aggregation inhibitor), cyclo-CEAGQLL (SEQ ID NO:39) (inactive on A β 42), or kanR (unrelated gene; negative control). Data shown are mean and s.e. of two biological replicates. Fold phage propagation is calculated as the number of phage generated from an infected culture divided by the number of phage (10^5) used to infect the culture.

[0040] FIG. 17. Evaluation of cyclic peptide sequences enriched by selection on hIAPP-T7n expressed from plasmid using the A β 42-GFP folding reporter. For comparison, positive (CKVWQLL (SEQ ID NO:30)) and negative (CEAGQLL (SEQ ID NO:39)) control cyclic peptides are included. Cyclic peptides are induced at 0.04 mM IPTG. Data reflects mean and s.e. of three biological replicates.

[0041] FIG. 18. Replicates of hIAPP aggregation assays with cyclo-CDLGVFR (SEQ ID NO:68) used to calculate t_{lag} values shown in FIG. 6D. Each graph reflects plots of individual values of three technical replicates. The hIAPP only data (0 μ M) in (a-d) is the same as in FIG. 7(a-d), respectively.

[0042] FIG. 19. Replicates of hIAPP aggregation assays with cyclo-CRCVFSG (SEQ ID NO:69) used to calculate t_{lag} values shown in FIG. 6F. Each graph reflects plots of individual values of three technical replicates. The hIAPP only data (0 μ M) in (a-d) is the same as in FIG. 7(a-d), respectively.

[0043] FIGS. 20A and 20B. Native mass spectra of 10 μ M hIAPP in the presence of 40 μ M cyclo-CDLGVFR (SEQ ID NO:68) (FIG. 20A) or cyclo-CRCVFSG (SEQ ID NO:69)

(FIG. 20B). Peak assignments are shown in blue with the theoretical masses given in parentheses. For the full spectra, see FIGS. 21A and 21B.

[0044] FIGS. 21A and 21B. Full native mass spectra of 10 μ M hIAPP in the presence of 40 μ M cyclo-CDLGVFR (SEQ ID NO:68) (FIG. 21A) or cyclo-CRCVFSG (SEQ ID NO:69) (FIG. 21B). Peak assignments are shown in blue with the theoretical masses given in parentheses.

[0045] FIG. 22. Sequence alignment of pIII-neg proteins and wild-type pIII. ClustalW2 was used to align all pIII-neg proteins against full-length pIII.

DETAILED DESCRIPTION OF THE INVENTION

[0046] One aspect of the invention is directed to selection systems.

[0047] The selection systems can comprise a host cell. The term “host cell” as used herein refers to a cell that can be infected by a library phage as described herein, replicate it, and package it into progeny phages that can infect fresh host cells. Exemplary activities of hosts include expression of genes of the phage, replication of the phage genome, and generation of progeny phage particles. One criterion to determine whether a cell is a suitable host cell for a given library phage is to determine whether the cell can support the viral life cycle of a wild-type viral genome from which the library phage is derived. For example, if the library phage is a modified M13 phage genome, as provided in some embodiments described herein, then a suitable host cell would be any cell that can support the wild-type M13 phage life cycle. Suitable host cells for various phages, such as for continuous evolution processes, are well known to those of skill in the art, and the disclosure is not limited in this respect.

[0048] The selection systems can further comprise a phage library comprising library phages configured to infect the host cell. “Phage library” as used herein refers to a collection of phages. The collection of phages can typically be contained and intermixed within a single medium or container. The term “phage” is used herein interchangeably with the term “bacteriophage” and refers to a virus that infects bacterial cells. Typically, phages comprise an outer protein capsid enclosing genetic material. The genetic material can be ssRNA, dsRNA, ssDNA, or dsDNA, in either linear or circular form. Phages and phage vectors are well known to those of skill in the art and non-limiting examples of phages that are useful for carrying out the methods provided herein are γ (Lysogen), T2, T4, T7, T12, R17, M13, MS2, G4, P1, P2, P4, Phi X174, N4, (6, and 029. In certain embodiments, the phage utilized in the present invention is M13. Such phages are preferably modified to have the characteristics described elsewhere herein. Additional suitable phages and host cells will be apparent to those of skill in the art, and the invention is not limited in this aspect. For an exemplary description of additional suitable phages and host cells, see Elizabeth Kutter and Alexander Sulakvelidze: *Bacteriophages: Biology and Applications*. CRC Press; 1st edition (December 2004), ISBN: 0849313368; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions (Methods in Molecular Biology)* Humana Press; 1st edition (December, 2008), ISBN: 1588296822; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages: Methods and Protocols, Volume 2: Molecular and Applied Aspects*

(*Methods in Molecular Biology*) Humana Press; 1st edition (December 2008), ISBN: 1603275649; all of which are incorporated herein in their entirety by reference for disclosure of suitable phages and host cells as well as methods and protocols for isolation, culture, and manipulation of such phages.

[0049] The library phages are preferably deficient in a replication gene that encodes a replication protein. “Replication protein” as used herein is a protein essential for the production of infectious progeny phages from the host cell when infected with the phage. “Deficient” in this context means that the phage does not have a replication gene that produces a sufficient amount and/or form of the replication protein to support a wild-type level of phage production. “Wild-type level of phage production” refers to a level of phage production from a phage that is equivalent to a library phage of the invention except that it contains a wild-type form of the replication gene. In some versions, the phage can lack a replication gene that expresses the replication protein, such that the replication protein is not expressed at all in the cell. In some versions, the phage can contain a modified form of a replication gene that expresses the replication protein, but the amount and/or a form of the replication protein is insufficient to support the wild-type level of phage production. “Phage production” in this context refers to the extracellular generation of infectious phage particles. The replication protein can be a protein involved in any part of the phage life cycle, including infection, phage genome replication, phage protein expression, phage assembly, phage release, and phage particle stability. Exemplary replication genes which the library phages can be deficient include gill, which encodes pIII; gIV, which encodes pIV; gVI, which encodes pVI; gVII, which encodes pVII; gVIII, which encodes pVIII; and gIX, which encodes pIX. Exemplary pIII sequences include the sequence encoded by the gill sequence provided in the following examples and sequences at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical thereto.

[0050] Each library phage preferably comprises a library gene configured to express a library peptide in the host cell. “Configured to express” as used herein with respect to a gene refers to a configuration of a gene such that it encodes a particular peptide and has the appropriate genetic elements (e.g., promoter, ribosomal binding site, etc.) to transcribe and translate the coding sequence into the peptide. The term “peptide” as used herein refers to a polymer of amino acid residues linked together by peptide bonds. The term “peptide” is used herein interchangeably with “polypeptide” and “protein.” The library peptide can be a peptide of any size, structure, or function. Typically, a library peptide will be at least three amino acids long. The library peptides can be fragments of naturally occurring proteins, entire naturally occurring proteins, modified (mutated) forms of naturally occurring proteins or fragments thereof, or any combination thereof.

[0051] It is preferred that at least a subset of the library peptides expressed by the library phages comprise different peptide sequences. The number of different peptides expressed by the library phages in a given phage library is referred to herein as the “library size.” In various versions of the invention, the library size can be greater than 10^1 , 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , or more. In various

versions of the invention, the library size can be up to 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} , 10^{13} , or more.

[0052] The library peptides can comprise linear peptides, cyclic peptides, or a combination thereof. In some versions of the invention, the library peptides comprise cyclic peptides. The cyclic peptides can have a head-to-tail configuration, a side-chain-to-side-chain configuration, or a side-chain-to-terminus configuration. Head-to-tail cyclic peptides have a configuration in which cyclization occurs exclusively through peptide bonds via the amino and carboxy groups of each constituent residue. Side-chain-to-side-chain cyclic peptides have a configuration in which cyclization occurs through side-chain bonding of constituent residue within the peptide. Side-chain-to-terminus cyclic peptides have a configuration in which either the amino or carboxy group of what would otherwise be an N-terminus or C-terminus, respectively, of a linear peptide bonds to a side chain of a constituent residue in the chain. Methods of configuring genes to produce cyclic peptides are known in the art and include, for example, the split-intein circular ligation of peptides and proteins (SICLOPPS) method¹⁵, among others. The library peptides can have any size capable of being made in a host cell. Exemplary sizes include from 2 amino acid residues to 250 amino acid residues or more, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, or any size between any two of the foregoing values.

[0053] The selection systems can further comprise a target fusion gene. The target fusion genes of the invention are genes configured to express a target fusion protein, particularly in the host cell. The target fusion protein can comprise a target protein fused to an RNA polymerase unit. The target protein and the RNA polymerase unit can be fused via a linker. The linker can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more amino acids long. The target fusion gene can be incorporated in the genome of the host cell or can be provided on a non-chromosomal vector (e.g., a plasmid) that can be introduced in the host cell, among other configurations. In some versions, the target fusion gene comprises a promoter that is activated by a transcription factor provided by the library phage, such that the library phage carries a gene configured to express the transcription factor. In some versions, this transcription factor is not produced by the host cell when not infected with the library phage, such that the host cell does not carry a gene configured to express the transcription factor when not infected with the phage.

[0054] The RNA polymerase unit can be an entire, functional RNA polymerase or subunit thereof. The RNA polymerase is preferably a DNA-dependent RNA polymerase.

[0055] The RNA polymerase in some versions is a promoter-specific RNA polymerase. “Promoter-specific RNA polymerase” refers to an RNA polymerase that specifically binds to a recognition sequence in a cognate promoter, wherein “cognate promoter” refers to a promoter comprising the recognition sequence. Non-limiting examples of promoter-specific RNA polymerases and their cognate promoters include the T7 RNA polymerase and the T7 promoter, the T3 RNA polymerase and the T3 promoter, and the SP6 RNA polymerase and the SP6 promoter, respectively. These and other RNA polymerases and their cognate promoters are well known in the art. Unless specified otherwise, reference to a specific RNA polymerase and cognate promoter (e.g.,

T7 RNA polymerase and T7 promoter, T3 RNA polymerase and T3 promoter, and SP6 RNA polymerase the SP6 promoter) encompasses the native forms of such elements as well as modified forms derived therefrom that maintain the same functionality. Examples of modified forms include sequence variants and split variants, the latter of which are described below. The RNA polymerase and its cognate promoter are preferably orthogonal to the host cell.

[0056] The RNA polymerase in some versions is a multi-subunit RNA polymerase. The multi-subunit RNA polymerase can be a natural multi-subunit RNA polymerase or a split RNA polymerase. Natural multi-subunit RNA polymerases are RNA polymerases that are comprised of multiple subunits in nature. Split RNA polymerases are RNA polymerases that are comprised of a single subunit in nature but can be recombinantly expressed as two or more separate subunits that combine to form a single, functional RNA polymerase. A number of split RNA polymerases are known in the art. Examples include split T7 RNA polymerases among others. See, e.g., Segall-Shapiro et al.⁵⁹ and Shis et al.⁶⁰ An exemplary split T7 RNA polymerase is provided in the following examples as the protein encoded by the “T7n” and “T7c” coding sequences. The exemplary split T7 RNA polymerase can accordingly be used in the present invention, as can split T7 RNA polymerases comprising subunits having at least 80%, least 85%, least 90%, least 95%, or least 99% sequence identity to the T7n subunit and/or the T7c subunit. RNA polymerases such as the T3 RNA polymerase and the SP6 RNA polymerase have high sequence homology to the T7 RNA polymerase and can be split in regions analogous to those in the split T7 RNA polymerases described herein. In preferred versions of the invention, the subunits of the multi-subunit RNA polymerase together exhibit RNA polymerase activity but do not individually exhibit RNA polymerase activity in isolation of each other.

[0057] Accordingly, in some versions of the invention, the RNA polymerase unit to which the target protein is fused in the target fusion protein is a first subunit of a multi-subunit RNA polymerase, such as a multi-subunit, promoter-specific RNA polymerase. In the case in which the subunit multi-subunit RNA polymerase is a split T7 RNA polymerase, the target protein can be fused either to the N-terminal subunit of the split RNA polymerase (preferably at the N-terminus thereof) or the C-terminal subunit of the split RNA polymerase (preferably at the C-terminus thereof). The N-terminal subunit of the split RNA polymerase is understood herein to constitute what would be the N-terminal portion of the unsplit RNA polymerase as it exists in nature. The C-terminal subunit of the split RNA polymerase is understood herein to constitute what would be the C-terminal portion of the unsplit RNA polymerase as it exists in nature. An exemplary N-terminal subunit is the peptide encoded by the T7n coding sequence provided in the following examples, or a peptide having at least 80%, least 85%, least 90%, least 95%, or least 99% sequence identity thereto. An exemplary C-terminal subunit is the peptide encoded by the T7c coding sequence provided in the following examples, or a peptide having at least 80%, least 85%, least 90%, least 95%, or least 99% sequence identity thereto.

[0058] In various versions, the RNA polymerase unit to which the target protein is fused in the target fusion protein consists of fewer than 250 amino acid residues, fewer than 225 amino acid residues, fewer than 200 amino acid residues, fewer than 175

amino acid residues, fewer than 150 amino acid residues, fewer than 100 amino acid residues, fewer than 75 amino acid residues, or fewer than 50 amino acid residues. In various versions, any linker linking the target protein and the RNA polymerase unit in the target fusion protein consists of fewer than 25 amino acid residues, fewer than 20 amino acid residues, fewer than 15 amino acid residues, fewer than 10 amino acid residues, fewer than 9 amino acid residues, fewer than 8 amino acid residues, fewer than 7 amino acid residues, fewer than 6 amino acid residues, fewer than 5 amino acid residues, fewer than 4 amino acid residues, fewer than 3 amino acid residues, or fewer than 2 amino acid residues.

[0059] In various versions of the invention, the target protein to which the RNA polymerase unit is fused in the target fusion protein can be any protein of interest suspected of having a structure and/or function affected, either directly or indirectly, by one or more of the library peptides. In some versions, the target protein binds, is predicted to bind, or is suspected of binding to at least one of the library peptides. In some versions, the target protein binds to at least one of the library peptides.

[0060] In some versions of the invention, the target protein is a protein that misfolds and/or aggregate and thereby causes the first subunit of the multi-subunit RNA polymerase to misfold and/or aggregate and thereby prevent binding to the one or more additional subunits of the RNA polymerase to result in a functional RNA polymerase. The library peptides in such cases can be peptides that promote, predicted to promote, or are suspected of promoting the folding of the target protein to thereby promote binding of the first subunit to the one or more additional subunits and thereby results in a functional RNA polymerase. The library peptides can promote the folding of the target protein, for example, by binding to it in the folded state to stabilize the folded structure.

[0061] In some versions of the invention, the target protein is an aggregation-prone protein. Aggregation-prone proteins are proteins that are intrinsically unstable and spontaneously unfold and aggregate under physiological conditions. Aggregation-prone proteins are well known in the art and are also referred to as “intrinsically disordered proteins.” Positive staining with Thioflavin T (2-[4-(Dimethylamino)phenyl]-3,6-dimethyl-1,3-benzothiazol-3-ium chloride), Thioflavin S (Product No. T1892, Sigma-Aldrich, St. Louis, MO), and Congo Red (sodium salt of 3,3'-([1,1'-biphenyl]-4,4'-diyl)bis(4-aminonaphthalene-1-sulfonic acid)) under physiological conditions are exemplary indicators of aggregation-prone proteins. Aggregation-prone proteins are involved in more than 50 human diseases. See, e.g., Iadanza et al. 2018⁶¹. Examples of aggregation-prone proteins include α -synuclein, amyloid beta (A β), microtubule-associated protein tau (τ), transthyretin (TTR), antibody light chains, fragments of immunoglobulin light chains, fragments of immunoglobulin heavy chains, full or N-term fragments of serum amyloid A protein (SAA), prion protein (PrP), β 2-microglobulin (β 2m), huntingtin exon 1 (HttEx1), ABri peptide, Adan peptide, N-term fragments of apolipoprotein A-I (ApoA1), C-term extended apolipoprotein A-II (ApoAII), apolipoprotein C-II (ApoCII), apolipoprotein C-III (ApoCIII), fragments of gelsolin, lysozyme (LYS), fragments of fibrinogen α -chain, N-term truncated cystatin C, islet amyloid polypeptide (IAPP), calcitonin, atrial natriuretic factor (ANF), N-term fragments of prolactin (PRL), insulin, medin, lactotransferrin, odontogenic ameloblast-

associated protein (ODAM), pulmonary surfactant-associated protein C (SP-C), galectin 7 (Gal-7), corneodesmosin (CDSN), C-term fragments of kerato-epithelin (β g-h3), semenogelin-1 (SGI), proteins S100A8/A, enfuvirtide among others (see Guthertz et al. 2022⁶² and Chiti et al. 2017¹).

[0062] In some versions of the invention, the target protein is an amyloidogenic protein. Amyloidogenic proteins are proteins that misfold and form amyloid aggregates. Amyloidogenic proteins are well-known in the art. See, e.g., Giasson et al.⁶³. Examples of amyloidogenic proteins include tau protein, alpha-synuclein, amyloid precursor protein, transthyretin, gelsolin, cystatin C, apolipoprotein A1, fibrinogen alpha chain, lactoferrin, beta-2 microglobulin, apolipoprotein A-II, semenogelin I, corneodesmosin, galectin-7, ITM2B, TGFBI, protein C, beta-lactoglobulin, serum amyloid P component, collagen (type XXV), alpha 1, and APPBP2, among others (see Guthertz et al. 2022⁶² and Chiti et al. 2017¹).

[0063] In versions employing a multi-subunit RNA polymerase, the selection system of the invention further preferably comprises one or more RNA-polymerase subunit genes configured to express one or more additional subunits of the RNA polymerase. In preferred versions of the invention, the first subunit and the one or more additional subunits are together sufficient to exhibit RNA polymerase activity. In preferred versions of the invention, the first subunit and the one or more additional subunits do not individually exhibit RNA polymerase activity in isolation of each other. The one or more RNA-polymerase subunit genes can be incorporated in the genome of the host cell or can be provided on one or more non-chromosomal vectors (e.g., plasmids) that can be introduced in the host cell, among other configurations.

[0064] The selection system of the invention further preferably comprises a selection gene. The selection gene can be incorporated in the genome of the host cell or can be provided on one or more non-chromosomal vectors (e.g., plasmids) that can be introduced in the host cell, among other configurations. The selection genes of the invention preferably comprise a cognate promoter of the RNA polymerase operationally connected to a coding sequence of a selection protein. The selection protein is a protein responsible for promoting either positive or negative selection of the library phage infecting a particular host cell, either by promoting replication of the library phage or by repressing its replication.

[0065] The selection protein in some versions comprises a replication protein, such as a replication protein in which the host cell is deficient. Use of a replication in this manner can promote positive selection of the infecting library phage. Exemplary replication proteins that can serve as suitable selection proteins include pIII, which is encoded by the coding sequence of gIII; pIV, which is encoded by the coding sequence of gIV; pVI, which is encoded by the coding sequence of gVI; pVII, which is encoded by the coding sequence of gVII; pVIII, which is encoded by the coding sequence of gVIII; and pIX, which is encoded by the coding sequence of gIX, among others.

[0066] The selection protein in some versions comprises a dominant-negative form of a replication protein, such as a dominant-negative form of a replication protein in which the host cell is deficient. Use of a replication in this manner can promote negative selection of the infecting library phage. Exemplary dominant-negative replication proteins that can

serve as suitable selection proteins include dominant-negative forms of pIII, pIV, pVI, pVII, pVIII, and pIX, among others. Examples of a dominant-negative pIII protein is the N-C83 variant, which is a pIII protein comprising an N-C83 domain, which has an internal deletion of 70 amino acids (i.e., amino acids 1-70) from the C-terminal domain of the pIII protein. The amino acid sequence of the N-C83 variant is shown in FIG. 5, which includes other examples of dominant-negative pIII variants. Other exemplary dominant-negative pIII sequences include the sequence encoded by the gIII-neg sequence provided in the following examples and sequences at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical thereto. Other examples of dominant-negative replication proteins include dominant-negative mutants of the pIV protein. Other examples of selection proteins that can be used for negative selection include other non-phage conventional counter selection genes.

[0067] In versions of the invention employing negative selection with a dominant-negative form of a replication protein as a selection protein, the selection system further preferably comprises a basal replication gene. The basal replication gene is preferably configured to provide minimal expression of a non-dominant-negative version of the replication protein. The basal replication gene can be configured as such by a coding sequence of the non-dominant-negative version of the replication protein being operationally connected to a low-copy number promoter, by being incorporated on a low-copy-number vector, or other configurations. The selection gene and the basal replication gene are preferably provided to the host cell via a vector in trans when negative selection is desired.

[0068] The selection systems of the invention can be used in methods of selection. The methods of selection can comprise a first selection. The first selection can comprise contacting a first population of host cells comprising multiple copies of the host cell with the phage library under conditions effective for the library phages to infect the host cells and thereby generate a first population of infected host cells, incubating the first population of infected host cells under conditions effective for production of first progeny phages to thereby produce a first selected phage library comprising the first progeny phages, and harvesting the first selected phage library. The first population of host cells can comprise or be introduced with various genes of the invention suitable for carrying out the selection, including the target fusion gene, the one or more RNA-polymerase subunit genes, and the selection gene. The introduction of these genes can occur simultaneously with the incubating or before or after the incubating. The introduction of these genes can be performed using any suitable method. Harvesting a phage library, such as the first selected phage library, can comprise collecting media in which the first progeny phages are produced. The harvesting can further comprise removing impurities (e.g., cells, cellular debris) through filtration, centrifugation, etc.

[0069] The methods of selection can further comprise a second selection. The second selection can comprise contacting a second population of host cells comprising multiple copies of the host cell with the first selected phage library under conditions effective for the first progeny phages to infect the host cells and thereby generate a second population of infected host cells, incubating the second population of infected host cells under conditions effective for produc-

tion of second progeny phages to thereby produce a second selected phage library comprising the second progeny phages, and harvesting the second selected phage library. The second population of host cells can comprise the same host cells as in the first population of host cells or different host cells. As in the first selection, the second population of host cells can comprise or be introduced with various genes of the invention suitable for carrying out the selection, including the target fusion gene, the one or more RNA-polymerase subunit genes, and the selection gene. The introduction of these genes can occur simultaneously with the incubating or before or after the incubating. The introduction of these genes can be performed using any suitable method.

[0070] The methods of selection can further comprise third selections, fourth selections, and so on, each comprising the same or similar steps as described above for the second selection.

[0071] In some versions, the first selection, the second selection, and/or any subsequent selections are positive selections. For example, the selection steps can be performed with a positive selection protein, such as a replication protein. In some versions, the first selection, the second selection, and/or any subsequent selections are negative selections. For example, the selection steps can be performed with a negative selection protein, such as a dominant-negative form of a replication protein or some other negative selection protein. In some versions, one or more positive selections are performed and then followed by one or more negative selections. In such versions, the target protein in the one or more negative selections is preferably different than the target protein employed in the one or more preceding positive selections.

[0072] Another aspect of the invention is directed to recombinant peptides. The recombinant peptides can comprise any library peptide of the invention. In some versions, the recombinant peptides of the invention comprise a library peptide of the invention that has been selected for reducing aggregation of an aggregation-prone protein using the selection methods of the invention.

[0073] In some versions, the recombinant peptides comprise cyclic peptides. The cyclic peptides in some versions comprise head-to-tail cyclic peptides comprising a sequence selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HVVGVIX_n (SEQ ID NO:3), HVHSYLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n (SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHXX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLVVDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14), NRLDV SX_n (SEQ ID NO:15), GLGHGNX_n (SEQ ID NO:16), RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), KVWQLAX_n (SEQ ID NO:19), RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVWGGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26), QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVYTAPX_n (SEQ ID NO:29), wherein X is any amino acid and n is any integer. In some versions, n is an integer from 0-30. In some versions, n is an integer such as 0, 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 or an integer between any two of

the foregoing values. In some versions, at least one X is selected from the group consisting of cysteine, serine, and threonine. In some versions, X is 1. In some versions, X is 1 and the X is selected from the group consisting of cysteine, serine, and threonine. In some versions, X is 1 and the X is cysteine. In some versions, the most N-terminal X of X_n is selected from the group consisting of cysteine, serine, and threonine. In some versions, the most N-terminal X of X_n is cysteine.

[0074] In some versions, the cyclic peptides of the invention are generated using a method referred to in the art as SICLOPPS (Valentine et al. 2018⁶⁴, Tavassoli 2017⁶⁵). SICLOPPS employs a cysteine, serine, or threonine as the first amino acid of an extein forming the final cyclic peptide (Valentine et al. 2018⁶⁴). Beyond this, there are no other limits on the number or identity of amino acids in the target peptide, allowing cyclic peptides of various sizes and sequences to be assembled (Tavassoli 2017⁶⁵). In some versions, the cyclic peptides of the invention are chemically synthesized, thereby removing the requirement for a cysteine, serine, or threonine as the first amino acid of the extein.

[0075] In some versions, the cyclic peptides can be isolated cyclic peptides.

[0076] The DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HVVGVIX_n (SEQ ID NO:3), HVHSYLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n (SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHXX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLWDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14), NRLDV SX_n (SEQ ID NO:15), and GLGHGNX_n (SEQ ID NO:16) cyclic peptides of the invention are particularly useful for reducing aggregation of human islet amyloid polypeptide. The RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), and KVWQLAX_n (SEQ ID NO:19), RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVWGGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26), QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVYTAPX_n (SEQ ID NO:29) cyclic peptides of the invention are particularly useful for reducing the aggregation of amyloid-β42.

[0077] Another aspect of the invention is directed to methods of reducing aggregation of an aggregation-prone protein. The methods can comprise contacting an aggregation-prone protein with a peptide of the invention. The contacting can be performed in vitro or in vivo. If performed in vivo, the peptide can be administered to a subject comprising the aggregation-prone protein using any suitable method. The peptide in such methods can comprise a library peptide that has been selected for reducing aggregation of the aggregation-prone protein using the selection methods of the invention. In some versions, the aggregation-prone protein comprises human islet amyloid polypeptide, and the cyclic peptide is selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HVVGVIX_n (SEQ ID NO:3), HVHSYLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n (SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHXX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLWDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14),

NRLDVXS_n (SEQ ID NO:15), and GLGHGNX_n (SEQ ID NO:16), wherein X is any amino acid and n is any integer. In some versions, the contacting the human islet amyloid polypeptide with the cyclic peptide is performed in a subject with type 2 diabetes.

[0078] In some versions, the aggregation-prone protein comprises amyloid-β42, and the cyclic peptide is selected from the group consisting of RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), and KVWQLAX_n (SEQ ID NO:19), RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVGWGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26), QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVYTAPX_n (SEQ ID NO:29), wherein X is any amino acid and n is any integer. In some versions, the contacting the amyloid-β42 with the cyclic peptide is performed in a subject with Alzheimer's disease.

[0079] "Gene" refers to a nucleic acid sequence capable of producing a gene product and may include such genetic elements as a coding sequence together with any other genetic elements required for transcription and/or translation of the coding sequence. Such genetic elements may include a promoter, an enhancer, and/or a ribosome binding site (RBS), among others. In some versions, multiple genes are configured in an operon, in which multiple coding sequences are operationally connected to a single promoter. Each coding sequence and promoter pair in such instances are considered herein to constitute separate genes, despite comprising the same promoter.

[0080] "Gene product" refers to products such as a polypeptide or an mRNA encoded and produced by a particular gene.

[0081] "Operationally connected" refers to a relationship between two genetic elements (e.g., a promoter and coding sequence), in which one of the genetic elements controls or affects the activity of the other genetic element.

[0082] "Endogenous" used in reference to a genetic element means that the genetic element is native to the cell in which it is disposed.

[0083] "Exogenous" used in reference to a genetic element means that the genetic element is not native to the cell in which it is disposed.

[0084] "Recombinant" as used herein with reference to nucleic acid molecules or polypeptides refers to nucleic acid molecules or polypeptides having a non-natural nucleic acid or polypeptide sequence, respectively. "Recombinant" as used herein with reference to a gene refers to a gene having a non-natural nucleic acid sequence, is exogenous, or is endogenous to a given cell but is disposed within the cell (e.g., within the cell's genome) at a locus different from the native form of the gene. "Recombinant" as used herein with reference to a cell refers to a cell that contains a recombinant nucleic acid molecule, polypeptide, or gene. Any gene, polypeptide, or protein described herein can be a recombinant gene, polypeptide, or protein.

[0085] A "homologous" gene or protein is a gene or protein inherited in two species from a common ancestor. While homologous genes or proteins can be similar in sequence, similar sequences are not necessarily homologous.

[0086] The terms "identical" or "percent identity", in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences

that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described herein (or other algorithms available to persons of skill) or by visual inspection. For sequence comparison and identity determination, one sequence typically acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence based on the designated program parameters. A typical reference sequence of the invention is any nucleic acid or amino acid sequence described herein. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see *Current Protocols in Molecular Biology*, F. M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2008)). One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity for purposes of defining homologs is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always>0) and N (penalty score for mismatching residues; always<0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an

expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915). In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. The above-described techniques are useful in determining sequence identity of sequences described herein.

[0087] Various methods for introducing genetic modifications are well known in the art and include homologous recombination, among other mechanisms. See, e.g., Green et al., *Molecular Cloning: A laboratory manual*, 4th ed., Cold Spring Harbor Laboratory Press (2012) and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press (2001).

[0088] The genes of the invention can be codon-optimized for the particular microorganism in which they are introduced. Codon optimization can be performed for any nucleic acid by a number of programs, including “GENEGPS”-brand expression optimization algorithm by DNA 2.0 (Menlo Park, CA), “GENEOPTIMIZER”-brand gene optimization software by Life Technologies (Grand Island, NY), and “OPTIMUMGENE”-brand gene design system by GenScript (Piscataway, NJ). Other codon optimization programs or services are well known and commercially available.

[0089] The term “introduce” used with reference to introducing a gene or other element into a cell refers to transferring the gene from outside of the cell to inside of the cell. Such introduction can be performed using any method in the art. Methods for introduction include but are not limited to protoplast fusion, transfection, transformation, conjugation, and transduction (see, e.g., Ferrari et al., Genetics, in Hardwood et al, (eds.), *Bacillus*, Plenum Publishing Corp., pp. 57-72, 1989).

[0090] The term “isolated” or “purified” means a material that is removed from its original environment, for example, the natural environment if it is naturally occurring, or a cultivation broth if it is produced in a recombinant host cell cultivation medium. A material is said to be “purified” when it is present in a particular composition in a higher concentration than the concentration that exists prior to the purification step(s).

[0091] U.S. Pat. Nos. 10,179,911 and 11,624,130 are incorporated herein by reference.

[0092] The elements and method steps described herein can be used in any combination whether explicitly described or not.

[0093] All combinations of method steps as used herein can be performed in any order, unless otherwise specified or clearly implied to the contrary by the context in which the referenced combination is made.

[0094] As used herein, the singular forms “a,” “an,” and “the” include plural referents unless the content clearly dictates otherwise.

[0095] Numerical ranges as used herein are intended to include every number and subset of numbers contained within that range, whether specifically disclosed or not. Further, these numerical ranges should be construed as providing support for a claim directed to any number or subset of numbers in that range. For example, a disclosure of from 1 to 10 should be construed as supporting a range of from 2 to 8, from 3 to 7, from 5 to 6, from 1 to 9, from 3.6 to 4.6, from 3.5 to 9.9, and so forth.

[0096] All patents, patent publications, and peer-reviewed publications (i.e., “references”) cited herein are expressly incorporated by reference to the same extent as if each individual reference were specifically and individually indicated as being incorporated by reference. In case of conflict between the present disclosure and the incorporated references, the present disclosure controls.

[0097] It is understood that the invention is not confined to the particular construction and arrangement of parts herein illustrated and described, but embraces such modified forms thereof as come within the scope of the claims.

EXAMPLES

Rapid Discovery of Cyclic Peptide Protein Aggregation Inhibitors by Continuous Selection

Summary

[0098] We report a new platform for the rapid phenotypic selection of protein aggregation inhibitors from genetically encoded cyclic peptide libraries in *E. coli* based on phage-assisted continuous evolution (PACE). Here, we developed a PACE-compatible selection for protein aggregation inhibition and employed it to identify cyclic peptides that suppress amyloid- β 42 (A β 42) and human islet amyloid polypeptide (hIAPP) aggregation. Additionally, we integrated a negative selection that removes false positives and off-target hits, significantly improving cyclic peptide selectivity. We show that selected inhibitors are active when chemically re-synthesized in in vitro assays. Our platform provides a powerful new approach for the rapid discovery of cyclic peptide inhibitors of protein aggregation and may serve as the basis for the future evolution of cyclic peptides with a broad spectrum of inhibitory activities.

Introduction

[0099] We have developed a new platform for selecting cyclic peptide protein aggregation inhibitors based on phage-assisted continuous evolution (PACE). By linking cyclic peptide activity to phage reproduction, we can perform 1-2 rounds of selection per day, leading to rapid discovery of active sequences using only standard molecular biology techniques and equipment. We demonstrate the utility of our system by using it to identify cyclic peptide sequences that inhibit the aggregation of disease-associated proteins amyloid- β 42 (A β 42) and human islet amyloid polypeptide (hIAPP). We also report the implementation of a negative selection that purges hits with off-target activity, resulting in selective cyclic peptide sequences. Finally, we show that our identified hIAPP inhibitors are active in in vitro amyloid formation assays. We envision that the reported system will improve the speed and convenience of cellular cyclic peptide selections and complement existing strategies for identifying inhibitors of disease-associated protein aggregation.

Results

[0100] We envisioned a general strategy for phage-assisted cyclic peptide discovery, shown in FIG. 1. We start with an selection phage (SP)-encoded cyclic peptide library generated by standard cloning techniques. This pool of SP is used to infect *E. coli* host cells with an accessory plasmid that links desired cyclic peptide activity with protein III (pIII) expression. pIII, which is encoded by gIII, is essential for both infection of and viral escape from the *E. coli* host cell. Upon infection, the *E. coli* biosynthesizes all SP-encoded genes, including the cyclic peptide. If the cyclic peptide possesses the activity under selection, it triggers pIII expression from the host cell, enabling phage propagation. Thus, only SP that carry active cyclic peptides enrich in the population.

Development of a PACE-Compatible Selection for Protein Aggregation Inhibitors

[0101] Previous strategies for selecting protein misfolding inhibitors in cells have utilized cell viability^{4,19,29,30} or signal from fluorescent protein fusions.^{31,32} These approaches cannot translate to PACE, which requires the selected activity to be coupled to gIII expression. Therefore, we turned to a split-T7 RNA polymerase (T7 RNAP) reporter used to evolve proteins with improved soluble expression.³³ In our new selection, host *E. coli* express an aggregation-prone protein (APP) fused to the N-terminal half of split-T7 RNAP (T7n). When the APP-T7n fusion is translated, it misfolds due to the APP, preventing T7n from associating with the C-terminal half of split-T7 RNAP (T7c; FIG. 2A). If a SP-encoded inhibitor binds to and prevents APP aggregation, the fused T7n will remain folded and join with T7c to produce full-length T7 RNAP. The reconstituted T7 RNAP then transcribes gIII, which we place under control of the T7 promoter (P_{T7}), producing pIII and allowing the SP to reproduce (FIGS. 2A and 2B).

[0102] We initially targeted amyloid- β 42 (A β 42) peptide, which is implicated in Alzheimer's Disease, as the APP. Our choice was motivated by the recent identification of a cyclic peptide inhibitor of A β 42 aggregation, cyclo-CKVWQLL^{31,32} (SEQ ID NO:30) that would provide a valuable control for optimizing and benchmarking our system. When fused to T7n, A β 42 should rapidly aggregate and prevent T7 RNAP reconstitution. Accordingly, this fusion produces low gIII transcription from P_{T7} , and SP carrying a control gene (kanR) reproduce poorly on host cells encoding the selection circuit (FIG. 2C). For comparison, we tested a fusion of a non-aggregating A β 42 mutant (F19S/L34P),³⁴ which gave rise to a ~500-fold increase in propagation of kanR SP (FIG. 2C). These results suggest that the split-T7 RNAP folding reporter can translate the degree of A β 42 aggregation to changes in gIII transcription.

Selection of SP-Encoded SICLOPPS Cyclic Peptides

[0103] Next, we determined if our selection could discriminate between SP encoding cyclic peptides that inhibit A β 42 aggregation and SP carrying inactive cyclic peptides. We used the SICLOPPS method to biosynthesize cyclic peptides that can be selected in *E. coli*.¹⁵ In SICLOPPS, a precursor fusion consisting of the C-terminal half of the Ssp DnaE split intein, the peptide sequence to be cyclized, and the N-terminal half of the same split intein undergoes intein splicing after translation to produce head-to-tail cyclized

peptides (FIG. 7). To benchmark our system, we used positive control sequence cyclo-CKVWQLL (SEQ ID NO:30), identified from a SICLOPPS cyclic peptide library to inhibit intracellular A β 42-GFP aggregation and shown to be active when chemically re-synthesized in both in vitro and *C. elegans* A β 42 aggregation models.³¹

[0104] We measured the reproduction of SP carrying SICLOPPS precursors for cyclo-CKVWQLL (SEQ ID NO:30) ("CKVWQLL SP") or a scrambled control cyclo-CVQWLKL (SEQ ID NO:31) ("CVQWLKL SP") on *E. coli* encoding the A β 42-T7n selection (FIG. 2B). CKVWQLL SP exhibited significantly increased (~500-fold) proliferation compared to SP carrying the scrambled control (FIG. 8A). However, we were concerned that the low overall phage amplifications observed (0.01- to 10-fold) could preclude selection of active cyclic peptides from a library. We thought that cyclic peptide maturation might not be fast enough to trigger pIII production during the SP life cycle (~15 min). Alternatively, SP backbone optimization could be required to coordinate the phage life cycle with our selection, which has been observed previously and can impose a prerequisite for implementing new PACE selections.^{27,35,36} To improve the overall fitness of cyclic peptide-encoding phage, we evolved the parent CKVWQLL SP on A β 42-T7n selection cells in PACE (FIG. 8B). We chose SP clone 2 (FIG. 8C) for downstream experiments. This evolved SP exhibited significantly increased overall proliferation compared to the parent (FIG. 8D) yet maintained excellent (~300-fold) discrimination between CKVWQLL SP and scrambled control (FIG. 2D). Installing point mutations at residues important for CKVWQLL (SEQ ID NO:30) activity³¹ produced decreases in SP propagation, further supporting dependence of phage fitness on cyclic peptide sequence (FIG. 2E).

[0105] Cyclic peptide formation by SICLOPPS occurs through intein splicing. We tested whether phage propagation was splicing dependent using CKVWQLL SP where the first cysteine in the N-terminal intein was mutated to alanine, preventing formation of the first splicing intermediate (FIG. 7). Unexpectedly, intein inactivation did not impact CKVWQLL SP propagation (FIG. 8D), suggesting that both intein-bound and mature cyclic peptide forms of this sequence are active. The lack of splicing dependence could not be explained by our evolved SP because the unevolved SP also exhibited splicing-independent activity (FIG. 8D). Additionally, it was not an artifact of our phage-based selection, as we also observed it in experiments testing plasmid-expressed SICLOPPS precursors on an A β 42-GFP folding reporter analogous to the one used to identify the CKVWQLL sequence (SEQ ID NO:30) (FIGS. 8E and 8F).³¹ This suggests that, when selecting SICLOPPS libraries, the species undergoing selection may exist as a mixture of pre-splice, splicing intermediate, and spliced products. However, SICLOPPS libraries have repeatedly produced cyclic peptides that function when chemically re-synthesized and tested in orthogonal assays,^{20,21,37,38} suggesting that this ambiguity does not prevent the discovery of active hits.

[0106] We performed a mock selection using a SP library where the four residues in CKVWQLL (SEQ ID NO:30) most critical for activity³¹ were diversified by saturation mutagenesis. We subjected the resulting CX₃QLX (X=amino acid encoded by NNK codons) library to PANCS-style selection²⁶ using host cells encoding the A β 42-T7n

selection circuit (FIG. 3A). After the first round, we split the SP pool into two separate lineages and selected them in parallel. Initially, we diluted the SP 2-fold per round and gradually increased the dilution factor to 10-fold in later rounds. After seven rounds, we observed a ~500-fold increase in SP pool fitness (measured by fold phage propagation on the selection cells), suggesting enrichment of active cyclic peptide sequences (FIG. 3B). Sanger sequencing of four clonal SP from each lineage revealed that the phage had either rediscovered CKVWQLL (SEQ ID NO:30) or enriched highly similar sequences (FIG. 3C). Together, these results suggest that we can select for active cyclic peptide sequences from an SP-encoded library using our system.

[0107] Next, we tested whether we could select active cyclic peptides from a library of randomized sequences by subjecting a cyclo-CX₆ SP library (X=NNK encoded amino acid) to PANCS²⁶ using host cells encoding the A β 42-T7n selection circuit (FIG. 3A). A total of ~3×10⁶ transformants was selected over seven rounds, over which we observed an increase in SP pool fitness (measured by fold phage propagation on the selection cells), suggesting enrichment of active cyclic peptide sequences (FIG. 3D). Deep sequencing revealed that the final SP pool had strongly converged, with the top sequence (CRVWCAR (SEQ ID NO:43)) accounting for ~65% of total sequencing reads (FIG. 3E). Alignment of the top 50 sequences (FIG. 9A) revealed high similarity to cyclo-CKVWQLL (SEQ ID NO:30), while the top sequence (CRVWCAR (SEQ ID NO:43)) bears strong resemblance to sequences highly enriched by the selection that identified cyclo-CKVWQLL (SEQ ID NO:30) in a prior study (e.g., CRVWCAL (SEQ ID NO:) and CRVWQAL (SEQ ID NO:)).²⁰

[0108] We measured the fitness of clonal phage encoding the top two sequences, CRVWCAR (SEQ ID NO:43) and CRVYQVL (SEQ ID NO:44), on the A β 42-T7n selection (FIG. 3F). SP carrying CRVWCAR (SEQ ID NO:43) exhibited robust propagation, comparable to that of CKVWQLL SP, while SP bearing CRVYQVL (SEQ ID NO:44) displayed more modest fitness. We also probed the sequence dependence of CRVWCAR (SEQ ID NO:43) through alanine substitutions (FIG. 3G). Mutation of residues 2-4 to alanine greatly diminished activity, consistent with the strong enrichment of R, V, and W at these positions by our selection (FIG. 9A). However, substitution of the first cysteine with alanine, which prevents SICLOPPS splicing, had no discernable effect on A β 42-T7n inhibition. This is consistent with our observations that CKVWQLL (SEQ ID NO:30) activity is also not splicing-dependent (FIGS. 8D-8F). Finally, we compared the activity of our sequences with CKVWQLL (SEQ ID NO:30) in the A β 42-GFP folding reporter assay (FIG. 3H). CRVWCAR (SEQ ID NO:43) had comparable activity to CKVWQLL (SEQ ID NO:30), while CRVYQVL (SEQ ID NO:44) appeared inactive. However, induction of the CRVYQVL (SEQ ID NO:44) construct greatly decreased cell growth (FIG. 9B), which may convolute rescue of A β 42-GFP fluorescence.

[0109] Together, these results suggest that our system can select for active cyclic peptide sequences from SP-encoded libraries.

Selection of Cyclic Peptide hIAPP Aggregation Inhibitors

[0110] Next, we examined if our system could discover inhibitors of other disease-associated APPs. We targeted human islet amyloid polypeptide (hIAPP), which forms

amyloid deposits in patients with type-2 diabetes that are hypothesized to contribute to R-cell dysfunction.⁴⁰ To create a selection for hIAPP aggregation inhibitors, we simply cloned hIAPP in place of A β 42 in our selection circuit (FIG. 2B). We used an SP-encoded CZ₆ SICLOPPS library (Z=NDT codon encoded amino acid: R, N, D, C, G, H, I, L, F, S, Y, or V) (~3×10⁶ total sequences). This codon set was chosen to reduce the theoretical library size and facilitate library cloning. The library (total of ~5×10⁶ transformants) was cloned on a SP obtained from an additional 250 hours of PACE using a more stringent selection (FIGS. 10A-10D).

[0111] We performed five rounds of selection, diluting the SP pool by ten-fold each round (FIG. 4A). There was a modest increase in SP pool fitness over the course of the selection (~35-fold). To verify that we selected phage with higher activity, we measured the propagation of the SP from each round on hIAPP selection cells, normalizing input phage titers to 105 pfu/mL to enable head-to-head comparison between phage pools. We observed a ~50-fold increase in SP pool fitness (FIG. 4B), suggesting enrichment of active cyclic peptides by the selection. Deep sequencing of the SP pool from round 5 revealed that a single sequence (CHWGV (SEQ ID NO:57)) accounted for ~95% of sequencing reads (FIGS. 4C and 4D), with the second most enriched sequence, CHVHSYL (SEQ ID NO:58), exhibiting some similarity to the top hit. Aside from these two sequences, we did not observe any clear trends in the remaining enriched cyclic peptides.

[0112] We synthesized both cyclo-CHWGV (SEQ ID NO:57) (FIG. 4D) and linear CHVVGVI (SEQ ID NO:67) and measured their effects on hIAPP aggregation in vitro. cyclo-CHWGV (SEQ ID NO:57) inhibited hIAPP aggregation in thioflavin T (ThT) fluorescence assays (FIG. 4E and FIG. 11), delaying the aggregation lag time (t_{lag}) in a dose-dependent fashion (FIG. 4F). In contrast, linear CHWGV (SEQ ID NO:67) showed greatly reduced ability to inhibit hIAPP (FIG. 4E), suggesting that cyclization is important for activity.

[0113] We further characterized the interaction of cyclo-CHVVGVI (SEQ ID NO:57) with hIAPP. TEM analysis corroborated the ability of cyclo-CHVVGVI (SEQ ID NO:57) to delay hIAPP aggregation onset, showing that no fibrils are observed when hIAPP is incubated with cyclo-CHVVGVI (SEQ ID NO:57) for 2 h (FIG. 4G and FIG. 12). In contrast, hIAPP produces numerous fibrils when incubated alone under the same conditions (FIG. 4H and FIG. 12). No gross differences in morphology were observed between fibrils formed by hIAPP alone or with cyclic peptide after incubation at longer times when all hIAPP should be aggregated (FIGS. 4I and 4J), suggesting that cyclo-CHVVGVI (SEQ ID NO:57) is not being incorporated into hIAPP fibrils or sequestering hIAPP in a different form of aggregate. Finally, analysis of the interaction of cyclo-CHWGV (SEQ ID NO:57) with hIAPP by native mass spectrometry primarily detected a 1:1 complex between cyclo-CHVVGVI (SEQ ID NO:57) and hIAPP (FIGS. 14A and 14B), suggesting that cyclo-CHVVGVI (SEQ ID NO:57) inhibits hIAPP aggregation through binding to monomeric hIAPP, consistent with the APP-T7n selection mechanism (FIG. 2A).

[0114] These results demonstrate that our system can be used to rapidly identify cyclic peptide inhibitors of a different aggregation-prone protein.

Negative Selection Identifies Selective Aggregation Inhibitors

[0115] The top sequences enriched by the hIAPP selection (CHWGV (SEQ ID NO:57) and CHVHSYL (SEQ ID NO:58)) resembled previously identified A β 42 aggregation inhibitors²⁰ and sequences from our own selections on A β 42-T7n. Because the amyloidogenic regions of A β 42 and hIAPP share sequence similarity,⁴² we wondered if our selection for hIAPP inhibitors discovered peptides that are also active on A β 42. Indeed, the hIAPP round 5 SP pool exhibited activity on both hIAPP-T7n and A β 42-T7n (FIG. 5A), suggesting that we had identified promiscuous sequences. However, the SP pools were inactive against a non-amyloidogenic target (an scFv that aggregates in *E. coli* cytosol). Non-specific hits and false positives are a concern for selection-based cyclic peptide discovery methods.^{14,43} Negative selection can increase selectivity and weed out false positives, but to our knowledge has not been explored in cellular cyclic peptide selections. These results motivated us to develop a negative selection that could quickly exclude cyclic peptides with non-selective activity or act outside of the APP in our selection circuit (for example, binding to and stabilizing T7n).

[0116] To create a negative selection, we placed gIII-neg, a dominant-negative form of gIII,⁴⁴ under P_{T7} in our selection circuit. Inhibition of a non-target APP (ex. A β 42) triggers pIII-neg expression and poisons phage proliferation. A second plasmid produces small amounts of pIII from the

round of negative selection, the SP pool fitness had increased by ~1000-fold, suggesting removal of promiscuous sequences (FIG. 5B). Comparison of the activity of the pre- and post-negative selection SP pools revealed a ~1000-fold decrease in A β 42-T7n aggregation inhibition, while the activity on hIAPP-T7n decreased modestly (<10-fold; FIG. 5C).

[0117] Deep sequencing of the post-negative selection pool (FIG. 5D and Table 1) revealed that CHVVGVI (SEQ ID NO:57) was absent from the top 10 sequences. Instead, >50% of sequencing reads mapped to CDLGVFR (SEQ ID NO:68) and CRCVSFG (SEQ ID NO:69) (FIG. 5D). These sequences were not highly enriched by positive selection (FIG. 4C), illustrating the ability of negative selection to produce large shifts in the cyclic peptide sequence distribution. We compared the activity of clonal SP encoding CHWGV (SEQ ID NO:57), CDLGVFR (SEQ ID NO:68), or CRCVSFG (SEQ ID NO:69) on selection cells encoding either hIAPP-T7n or A β 42-T7n (FIG. 5E). While CHWGV SP propagated robustly on both targets, CRCVSFG SP were more selective, with ~10-fold higher activity on hIAPP vs. A β 42 selection cells. SP encoding CDLGVFR (SEQ ID NO:68) were strikingly selective, exhibiting ~1000-fold greater propagation on hIAPP vs. A β 42. We observed similar trends with the A β 42-GFP folding reporter when CHWGV (SEQ ID NO:57), CDLGVFR (SEQ ID NO:68), and CRCVSFG (SEQ ID NO:69) were expressed from plasmid (FIG. 17), suggesting that selectivity is not an artifact of our phage system.

TABLE 1

Comparison of top 10 cyclic peptide sequences enriched from positive selection on hIAPP or subsequent negative selection against A β 42 as determined by high-throughput sequencing.

Sequence	Rank (positive selection)	Percent total reads (positive selection)	Rank (negative selection)	Percent total reads (negative selection)	Fold enrichment in negative vs. positive selection
CHVVGVI (SEQ ID NO: 57)	1	94.40	20	0.29	0.003
CHVHSYL (SEQ ID NO: 58)	2	1.82	4	6.03	3.31
CLNYFHG (SEQ ID NO: 59)	3	0.35	3	7.05	20.08
CYILSIG (SEQ ID NO: 60)	4	0.34	18	0.37	1.10
CCGLYNI (SEQ ID NO: 61)	5	0.31	31	0.15	0.48
CCHSFFR (SEQ ID NO: 62)	6	0.29	12	0.83	2.87
CGIRSLG (SEQ ID NO: 63)	7	0.22	7	4.48	20.09
CISCHYG (SEQ ID NO: 64)	8	0.13	34	0.12	0.92
CIYFHHH (SEQ ID NO: 65)	9	0.12	8	3.33	28.44
CVSYILL (SEQ ID NO: 66)	10	0.11	16	0.45	4.21
CDLGVFR (SEQ ID NO: 68)	14	0.08	1	31.22	410.79
CRCVSFG (SEQ ID NO: 69)	20	0.04	2	24.55	701.46
CFNLVVD (SEQ ID NO: 72)	15	0.08	5	4.85	64.65
CFFRGSD (SEQ ID NO: 73)	18	0.06	6	4.73	81.53
CNRLDVS (SEQ ID NO: 76)	26	0.02	9	1.84	83.68
CGLGHGN (SEQ ID NO: 77)	53	0.009	10	1.12	124.22

phage shock promoter (FIG. 16A). All SPs infecting cells encoding this negative selection receive sufficient pIII to propagate. However, SP that carry a promiscuous or false positive cyclic peptide will also produce pIII-neg and generate non-infectious progeny (FIGS. 16B and 16C). We subjected the hIAPP round 5 phage population to four rounds of negative selection, diluting the phage by 100-fold each round (FIG. 5B). The starting fitness of the hIAPP SP pool on the negative selection was low, confirming presence of library members active on A β 42. However, by the fourth

[0118] We examined the ability of chemically synthesized cyclo-CDLGVFR (SEQ ID NO:68) and cyclo-CRCVSFG (SEQ ID NO:69) (FIGS. 6A and 6B) to inhibit in vitro hNAPP aggregation. Both cyclic peptides inhibited hNAPP fibrillation and extended t_{lag} (FIGS. 6C-6F and FIGS. 18 and 19). In contrast, linear CDLGVFR (SEQ ID NO:78) was completely inactive (FIG. 6C), while linear CRCVSFG (SEQ ID NO:79) moderately inhibited hNAPP aggregation, but with reduced activity compared to cyclo-CRCVSFG (SEQ ID NO:69) (FIG. 6E). Like cyclo-CHWGV (SEQ ID

NO:57), cyclo-CDLGVFR (SEQ ID NO:68) and cyclo-CRCVSFG (SEQ ID NO:69) exhibited similar activity to MCIP-2a at lower concentrations, but were less active than MCIP-2a at higher peptide:hIAPP stoichiometries. TEM analysis confirmed that both cyclic peptides delay hNAPP aggregation (FIGS. 6G and 6H and FIG. 12) and do not change the overall appearance of hNAPP fibrils (FIGS. 6I and 6J). Additionally, we observed 1:1 complexes of both cyclo-CDLGVFR (SEQ ID NO:68) and cyclo-CRCVSFG (SEQ ID NO:69) with hNAPP by native mass spectrometry (FIGS. 20A and 20B), suggesting that, like cyclo-CHWGV (SEQ ID NO:57), these cyclic peptides are primarily hNAPP monomer binders.

[0119] Together, these results confirmed that our negative selection had purged promiscuous sequences to successfully uncover active and selective cyclic peptide inhibitors of hNAPP aggregation.

Discussion

[0120] We developed a new platform for identifying cyclic peptide protein aggregation inhibitors that leverages elements of PACE to increase the speed and convenience of selection. This method does not require specialized equipment (such as a cell sorter) and, aside from the initial generation of the SP library, eliminates the need for additional cloning or transformation steps. As a demonstration of the platform's utility, we applied it to identify new cyclic peptide inhibitors of hIAPP aggregation in less than a week. In addition, the observation that enriched inhibitors also exhibited activity on a different aggregation-prone protein, A β 42, prompted us to create a negative selection strategy to remove promiscuous sequences. We used this negative selection to identify cyclic peptides that selectively inhibited hIAPP. Off-target activity or false positives can plague selections and high-throughput screens of chemical libraries for active compounds, resulting in extra time and labor expended on isolating desired hits. Thus, negative selections such as the one used here can improve hit quality by purging undesired library members and streamline the isolation of desirable cyclic peptide sequences.

[0121] To our knowledge, the cyclic peptide hIAPP inhibitors we report are the first to be identified through a selection-based method. Previously, macrocyclic peptides have been generated by rational design to inhibit hIAPP aggregation through displayed hIAPP sequence mimics^{41,45} or aromatic moieties.^{46,47} Our cyclic peptides bear little sequence resemblance to rationally designed inhibitors, suggesting that unbiased selection can uncover new starting points for inhibitor development. Future efforts may improve the potency of hIAPP inhibitors identified by our platform through developing more stringent selections that require more active cyclic peptides to pass.

[0122] An unexpected observation made here is that the SICLOPPS cyclic peptide sequences from this work and other studies²⁰ may not require intein splicing to inhibit target protein aggregation. Because splicing efficiency varies by extein sequence,⁴⁹ it is difficult to predict the ratio of unspliced to spliced product for each cyclic peptide library member. Thus, the species under selection may be an intein-bound or spliced form of the peptide, or both. Our hits are active as chemically synthesized cyclic peptides and exhibit diminished activity when employed in linear form, suggesting that a cyclic conformation is important for activity in our selection for hIAPP inhibitors. However, other targets could behave differently, and moving forward, it may be preferable to switch the Ssp intein in SICLOPPS with with faster splicing homologs to minimize the time cyclic peptide sequences undergoing selection spend in the intein-bound form.^{49,50}

[0123] One limitation of our platform is the size of the cyclic peptide libraries employed, which is restricted by cloning efficiencies. This constraint is not unique to our system and applies to other cellular selections.^{13,16} A potential route to increasing library size is to further mutate the SP using host cell-encoded mutagenesis plasmids.³⁹

[0124] Although here we exclusively select for protein aggregation inhibitors, in principle our platform could be used to select for any cyclic peptide activity that can be linked to a genetic selection. PACE selections have been developed for a wide range of biologically significant activities, such as protein-protein interaction,^{28,51} protein-DNA interaction,^{35,52,53} and ternary complex formation²⁶ to name a few, and these selections could be adapted for use in cyclic peptide discovery. Our system may also be able to accommodate other types of genetically encoded peptide libraries.^{16,18} Thus, we anticipate that our work will provide a versatile and accessible new option for cyclic peptide discovery.

Materials and Methods

General Methods

[0125] Antibiotics (Gold Biotechnology) were used at the following working concentrations: ampicillin, 50 μ g/L; spectinomycin, 100 μ g/mL; chloramphenicol, 25 μ g/mL; kanamycin, 50 μ g/mL; tetracycline, 10 μ g/mL; streptomycin, 50 μ g/mL. HyClone water (GE Healthcare Life Sciences) was used for PCR reactions and cloning. For all other experiments, water was purified using a MilliQ purification system (Millipore). Phusion U Hot Start DNA polymerase (Thermo Fisher Scientific) or Q5 polymerase (New England Biolabs) were used for PCRs. A full list of plasmids used in this work is given in Table 2. Key primers are provided in Table 3.

TABLE 2

Plasmids and selection phages used in this work.								
Plasmid			ORF1			ORF2		
Name	Resistance	Origin	promoter	RBS	gene(s)	promoter	RBS	gene
MP6 ³⁹	chlor	cloDF13	P _{BAD}	native dnaQ	dnaQ926, dam, seqA, emrR, PBS2 UGI, pmCDA1			
pLY009	spec	ColE1	P _{psp}	SD8 ⁵⁶	Aβ42-T7n			
pTW006ap1a ³³	amp	SC101	P _{T7a}	SD8	gIII (recoded), luxAB	P _{pro1} ⁵⁷	SD8	T7c

TABLE 2-continued

Plasmids and selection phages used in this work.								
Plasmid			ORF1			ORF2		
Name	Resistance	Origin	promoter	RBS	gene(s)	promoter	RBS	gene
pTW357b	spec	ColE1	P _{T7a}	sd5 ⁵⁶	gIII (recoded)	P _{pro1} ⁵⁷	SD8	T7c R632S
pTW357d5	spec	ColE1	P _{T7a}	sd5	gIII (recoded)	P _{pro5} ⁵⁷	SD8	T7c R632S/ Q649S
pTW357bN	spec	ColE1	P _{T7a}	sd5	gIII-neg	P _{pro1} ⁵⁷	SD8	T7c R632S
pTW358b	kan	p15A	P _{psp}	SD8	Aβ42-T7n			
pTW358c	kan	p15A	P _{psp}	SD8	Aβ42 F19S/L34P-T7n			
pTW358f	kan	p15A	P _{psp}	SD8	hIAPP-T7n			
pTW358d	kan	p15A	P _{psp}	SD8	6xHis-scFv-C11-T7n			
pLY034b	amp	pSC101	P _{psp}	sd5	gIII (recoded)			
pBL066c	spec	ColE1	P _{tet}	SD8	Aβ42-GFP			
pLY002tac	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (parent)			
pLY002tacB	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (parent)			
pLY002tacD	kan	p15A	P _{tac}	SD8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (parent)			
pLY063a	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 1 clone 2)			
pLY064D	kan	p15A	P _{tac}	SD8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 1 clone 2)			
pLY063B	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 1 clone 2)			
pLY090a	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CRVWCAR (SEQ ID NO: 43) (PACE 1 clone 2)			
pLY090b	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CRVYQVL (SEQ ID NO: 44) (PACE 1 clone 2)			
pLY071	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 1)			
pLY071B	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 2 clone 1)			
pLY071D	kan	p15A	P _{tac}	SD8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 1)			
pLY071c	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CHVVGVI (SEQ ID NO: 57) (PACE 2 clone 1)			
pLY071e	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CDLGVFR (SEQ ID NO: 68) (PACE 2 clone 1)			
pLY071f	kan	p15A	P _{tac}	SD8	SICLOPPS precursor for cyclo-CRCVFSG (SEQ ID NO: 69) (PACE 2 clone 1)			
spLY001		M13	P _{gIII}	sd8 ⁵⁶	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (parent)			
spLY001d		M13	P _{gIII}	sd8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (parent)			
spLY002		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CVQWLKL (SEQ ID NO: 31) (parent)			

TABLE 2-continued

Plasmids and selection phages used in this work.								
Plasmid			ORF1			ORF2		
Name	Resistance	Origin	promoter	RBS	gene(s)	promoter	RBS	gene
spLY001M7		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (parent)			
spLY014		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 1 clone 2)			
spLY014D		M13	P _{gIII}	sd8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 1 clone 2)			
spLY014b		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 1 clone 2)			
spLY014c		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CVQWLKL (SEQ ID NO: 31) (PACE 1 clone 2)			
spLY014d		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEVWQLL (SEQ ID NO: 33) (PACE 1 clone 2)			
spLY014e		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKAWQLL (SEQ ID NO: 34) (PACE 1 clone 2)			
spLY014f		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKVGQLL (SEQ ID NO: 35) (PACE 1 clone 2)			
spLY014g		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAWQLL (SEQ ID NO: 36) (PACE 1 clone 2)			
spLY014h		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEVGQLL (SEQ ID NO: 37) (PACE 1 clone 2)			
spLY014i		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKAGQLL (SEQ ID NO: 38) (PACE 1 clone 2)			
spLY032		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CRVWCAR (SEQ ID NO: 43)			
spLY033		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CRVYQVL (SEQ ID NO: 44)			
spLY032a		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-ARVWCAR (SEQ ID NO: 53)			
spLY032b		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CAVWCAR (SEQ ID NO: 54)			
spLY032c		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CRAWCAR (SEQ ID NO: 55)			
spLY032d		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CRVACAR (SEQ ID NO: 56)			
spLY006		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 1 with 6xHis tag)			
spLY006b		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 2 clone 1 with 6xHis tag)			
spLY017		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 1)			
spLY017b		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 2 clone 1)			
spLY017D		M13	P _{gIII}	sd8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 1)			

TABLE 2-continued

Plasmids and selection phages used in this work.								
Plasmid			ORF1			ORF2		
Name	Resistance	Origin	promoter	RBS	gene(s)	promoter	RBS	gene
spLY018		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 2)			
spLY018b		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 2 clone 2)			
spLY018D		M13	P _{gIII}	sd8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 2)			
spLY019		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 3)			
spLY019b		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 2 clone 3)			
spLY019D		M13	P _{gIII}	sd8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 3)			
spLY020		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 4)			
spLY020b		M13	P _{gIII}	sd8	SICLOPPS precursor for cyclo-CEAGQLL (SEQ ID NO: 39) (PACE 2 clone 4)			
spLY020D		M13	P _{gIII}	sd8	splicing-deficient SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30) (PACE 2 clone 4)			
spLY021		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CHVVGVI (SEQ ID NO: 57)			
spLY026a		M13	PgIII	sd8	SICLOPPS precursor for cyclo-AHVVGVI (SEQ ID NO: 80)			
spLY026b		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CAVVGVI (SEQ ID NO: 81)			
spLY026c		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CHAVGVI (SEQ ID NO: 82)			
spLY026d		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CHVAGVI (SEQ ID NO: 83)			
spLY026e		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CHVVAVI (SEQ ID NO: 84)			
spLY026f		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CHVVGAI (SEQ ID NO: 85)			
spLY026g		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CHVVGVA (SEQ ID NO: 86)			
spLY026l		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CLDCVFG (SEQ ID NO: 87)			
spLY026m		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CHLDIGC (SEQ ID NO: 88)			
spLY026n		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CLSRRCG (SEQ ID NO: 89)			
spLY022a		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CRCVFSG (SEQ ID NO: 69)			
spLY022b		M13	PgIII	sd8	SICLOPPS precursor for cyclo-CDLGVFR (SEQ ID NO: 68)			

TABLE 3

Key primers used in this work.		
Primer Name	Sequence (5' to 3')	Purpose
LY0003	acagtttgUggggcgatcgc (SEQ ID NO: 90)	SP cyclic peptide library cloning (reverse primer for USER cloning)
LY0106	acaactgUnnknnknnknnknnknnktgtctcag tttcggcaccg (SEQ ID NO: 91)	SP CXXQLX cyclic peptide library cloning (forward primer for USER cloning)
LY0107	acaactgUndtndtndtndtndtndtndtgtctcag tttcggcaccg (SEQ ID NO: 92)	SP CZ ₆ cyclic peptide library cloning (forward primer for USER cloning)
LY0145	ggagtgcaaagaatatttgatattgg (SEQ ID NO: 93)	Amplify cyclic peptide sequence for HTS
LY0146	caatggggccgtactcaac (SEQ ID NO: 94)	Amplify cyclic peptide sequence for HTS
LY0147	acactctttccctacacgacgctcttcgatctg gagtgcaaagaatatttgatattgg (SEQ ID NO: 95)	adapter primer for HTS
LY0148	gtgactggagttcagacgtgtgctcttcgatct caatggggccgtactcaac (SEQ ID NO: 96)	adapter primer for HTS

[0126] Strain S206035 was used for plasmid cloning, amplification, and phage assays. Strain S220835 was used for SP cloning and plaque assays. Plasmids and SPs were cloned by USER assembly or blunt-end ligation. Competent cells were prepared and transformed using the TSS method. Phage propagation and plaque assays were performed as previously described.⁵⁴ Unless otherwise noted, phage propagation assays used an input of 105 phage to infect a 2 mL culture of host cells.

Phage-Assisted Continuous Evolution

[0127] In general, PACE was set up and run as previously described.²⁷

[0128] PACE 1. S2060s transformed with pLY009, pTW6ap1a, and MP6 were maintained in a 40 mL chemostat. Lagoons (15 mL each) were infected with evolved spLY001 from PACE at an initial titer of 4×10^4 pfu/mL and maintained at a flow rate of 0.5 V/h. Lagoon flow rates were increased to 1 V/h at 19 h, decreased to 0.8 V/h at 66 h, increased back to 1 V/h at 159.5 h, further increased to 1.5 V/h at 257 h, and finally to 2 V/h at 305 h. The experiment ended at 351 h.

[0129] PACE 2. S2060s transformed with pTW357d5, pTW358b, and MP6 were maintained in a 40 mL chemostat. Lagoons (15 mL each) were infected with final phage populations collected from PACE 1 at an initial titer of 1×10^5 pfu/mL and maintained at a flow rate of 0.5 V/h. Lagoon flow rates were increased to 1 V/h at 42.5 h, 1.5 V/h at 117 h, followed by 2 V/h at 160.5 h then 2.5 h at 189 h, and finally 3 V/h at 208 h. The experiment ended at 254.5 h.

A β 42-GFP Fluorescence Assay

[0130] Single colonies of S2060s transformed with pBL066c and SICLOPPS plasmid were used to make overnight cultures in 2 $\times$ YT with maintenance antibiotics. The

overnight culture was diluted (1:100) into 1 mL DRM on a 96 deep-well plate (VWR, 75870-796) with maintenance antibiotics and IPTG (0, 0.04, 0.2, or 1 mM). The diluted culture was incubated at 37° C. with shaking for 2.5 h, after which aTc (200 ng/mL) was added. 100 μ L of the culture was transferred onto a black clear-bottom 96-well plate (VWR, 89131-680) to measure fluorescence signal 2, 3, 5 h after addition of aTc. GFP fluorescence signal (excitation wavelength=485 nm, emission wavelength=535 nm) was then measured using a Tecan M Plex microplate reader.

Cyclic Peptide Library Cloning

[0131] The reverse primer LY0003 and the degenerate forward primers LY0106 (CX₆) or LY0107 (CZ₆) (Table 3) were used to clone the phage-encoded cyclic peptide libraries with spLY006b as the template. The resulting PCR product was purified by PCR clean-up and concentrated to 100 ng/ μ L. The purified PCR product was assembled by USER assembly. The USER assembled product was purified again by PCR clean-up and diluted to 50 ng/ μ L. For each aliquot, 250 ng of assembled DNA was transformed into chemicompetent S2208 cells. After heat shock, the aliquots were immediately combined into pre-warmed (37° C.) outgrowth media (2 $\times$ YT containing 3 mM glucose, 500 μ L for each aliquot) and incubated at 37° C. with shaking at 250 rpm for 1 h. After outgrowth, the culture was centrifuged at 8,000 rcf for 2 min and the supernatant containing the phage library was collected. This process resulted in 3×10^6 and 5×10^6 independent clonal phage for the CX₆ and CZ₆ libraries, respectively, as measured by plaque assay.

Cyclic Peptide Library Selection

[0132] The phage libraries obtained immediately after cloning were first amplified to introduce degeneracy: mid-log S2208 cells grown in Davis Rich Media (DRM) with maintenance antibiotics at 37° C. with shaking were infected

with the phage library and phage propagation allowed to occur for 6 h before the culture was centrifuged at 8,000 rcf for 2 min and the supernatant containing the expanded phage library collected. The expanded phage libraries contained ~1011 total phage as measured by plaque assay.

[0133] Single colonies of host cells transformed with the appropriate selection plasmids were used to make overnight cultures in 2×YT with maintenance antibiotics. The overnight culture was diluted (1:100) into 4-5 mL DRM with maintenance antibiotics and incubated at 37° C. with shaking until the OD₆₀₀ of the culture reached ~0.5, upon which the expanded phage library was added at a titer of 1×10⁶ pfu/mL. The culture was incubated at 37° C. with shaking overnight (13-18 h). The next day, the culture was centrifuged at 8,000 rcf for 2 min to obtain the supernatant containing phage. The collected phage was diluted into fresh host cell cultures (10 to 1,000-fold) for the next round of selection, while the rest was stored at 4° C. for downstream analysis. This process was repeated until a significant increase in phage propagation was observed compared with the parent library. Both positive and negative selection rounds were performed using this procedure.

High-Throughput Sequencing and Data Analysis

[0134] Sample preparation. Phage stocks collected from the final rounds of library selection were first amplified using LY0145 and LY0146. The PCR product was purified, and adapters added using LY0147 and LY0148. The resulting PCR product was purified and barcoded through a final PCR using index primers to add unique indices for shared Illumina NovaSeq sequencing. The final PCR product was purified, and the concentration measured by Nanodrop. Further sample quality control was performed by the University of Wisconsin-Madison Biotechnology Center (UWBC).

[0135] High-throughput sequencing and data analysis. High-throughput sequencing was performed by UWBC using a shared sequencing service on the NovaSeq 6000 platform. Sequencing reads were demultiplexed by UWBC and analyzed using custom-written Python scripts.

Solid-Phase Peptide Synthesis

[0136] General procedure for solid-phase peptide synthesis (SPPS). (see herein for abbreviations) Dawson Dbz AM resin (Sigma-Aldrich) (for head-to-tail cyclized peptides) or Wang resin (CEM) (for linear peptides) was added to an empty SPE cartridge with pre-inserted frit (Agilent, 6 mL capacity). The resin was suspended in 2 mL DCM with shaking and allowed to swell for 40 min at RT. DCM was drained and the resin washed 3 times with 2 mL DCM, followed by 3 times with 2 mL DMF. The resin was deprotected by adding freshly prepared 20% piperidine in DMF and shaking for 30 min, followed by three washes with 2 mL DMF. Couplings were performed with 4 eq. Fmoc-protected L-amino acid, 3.9 eq. HATU, and 8 eq. DIPEA in 2 mL DMF for 1 h with agitation. After coupling, the resin was washed with 2 mL DMF three times. The deprotection and coupling steps were repeated until the peptide was completed. Couplings for Fmoc-L-Arg(Pbf)-OH was repeated once. For head-to-tail cyclic peptides, Boc-L-Cys(Trt)-OH was used as the final amino acid. Upon completion, the resin was washed 3 times with 2 mL DMF and 3

times with 2 mL DCM. Note: the resin was dried by N₂ flow after every washing step in this procedure.

[0137] Dbz linker activation. For head-to-tail cyclized peptides, 4-nitrophenyl chloroformate (4 eq) was dissolved in 2 mL DCM and added to the resin. After shaking for 30 min, the solution was drained, and the resin washed 3× with 2 mL DCM. The 4-nitrophenyl chloroformate addition and washing steps were repeated once. The resin was then washed 3× with 2 mL DMF. A solution of 110 µL DIPEA in 2 mL DMF was added and the resin agitated for 10 min, followed by draining the solution. This step was repeated twice. The resin was then washed 3× with 2 mL DMF, 3× with 2 mL DCM, and 3× with 2 mL diethyl ether, then allowed to dry under a flow of N₂. Note: the resin was dried by N₂ flow after every washing step in this procedure.

[0138] Deprotection and cleavage. The resin was transferred to a 15 mL conical tube, then treated with cleavage solution (2 mL 90% TFA, 5% DCM, 2.5% H₂O, and 2.5% TIPS) for 2 h with agitation. The solution was filtered into a 50 mL conical tube through an empty syringe plugged with cotton. The remaining resin was washed with another 2 mL cleavage solution and filtered combined into the 50 mL conical in the same manner. 40 mL diethyl ether was added into the conical, which was then incubated at -20° C. for 1 h to facilitate precipitation of peptide product. After precipitation, the solution was centrifuged at 3,000 rpm for 10 min at 4° C. The supernatant was decanted, and the precipitate was dried under N₂ flow to evaporate remaining volatiles. The dried precipitate was then dissolved in 5 mL 20% aq. MeCN, frozen on dry ice for 1 h, and lyophilized.

[0139] Linear peptide purification. Lyophilized linear peptide was dissolved in 2 mL 20% aq. MeCN. The solution was passed through a 0.22 µm syringe filter into a fresh 15 mL conical tube. Another 2 mL 20% aq. MeCN was used to wash the old conical and filtered into the new conical tube. All 4 mL of peptide-containing solution was purified by HPLC (Shimadzu, CBM-20A, LC-20AP, SPD-20AV, FRC-10A) in a single injection onto a preparative C18 column (Shimadzu, Premier Elite Polar 10 µ 150×30 mm). HPLC conditions: Flow rate, 25 mL/min. Mobile phase A: H₂O containing 0.1% TFA. Mobile phase B: MeCN. A linear gradient of 10% to 50% mobile phase B over 32 min was used to purify the Nbz-containing linear peptide. Fractions were analyzed by MALDI-TOF to confirm mass of the Nbz-containing linear peptide, then combined and lyophilized.

[0140] Cyclization of Nbz-containing linear peptides. Lyophilized linear peptide was dissolved in 4 mL cyclization buffer (0.1 M Na₂HPO₄, 6 M guanidinium chloride, and 20% v/v MeCN in H₂O; pH 6.8-7.2) and incubated at 50° C. with rotation for 2-4 h.

[0141] Cyclic peptide purification. After cyclization, the solution was directly injected onto a semi-preparative C18 column (Kromasil, Eternity-5-C18 10×250 mm) HPLC conditions: flow rate, 5 mL/min. Mobile phase A: H₂O containing 0.1% TFA. Mobile phase B: MeCN containing 0.1% TFA. A linear gradient of 20% to 45% mobile phase B over 32 min was used to purify the cyclic peptide. Fractions were analyzed with MALDI-TOF to confirm desired mass of the cyclic peptide, then combined and lyophilized. After lyophilization, H₂O was added to dissolve the cyclic peptide to a stock concentration of 1 mM. The prepared 1 mM cyclic peptide stock solution was quickly aliquoted into low-protein-binding Eppendorf tubes, frozen on dry ice for 1 h

and lyophilized to dried peptide solid. Peptides were freshly dissolved before use. Additional freeze-thaw cycles were kept to a minimum to prevent degradation of cyclic peptide stocks.

[0142] Cyclic peptide characterization. Analytical HPLC was performed using a HPLC system (Shimadzu, DGU-20A5R, LC-20AT, SIL-10AF, SPD-M20A, CTO-20A) equipped with an analytical C18 column (Kromasil, Eternity-5-C18 4.6×250 mm). HPLC conditions: at a flow rate of 1 mL/min. A binary solvent system with 90% mobile phase A (M.Q. H₂O, 0.1% TFA) and 10% mobile phase B (ACN, 0.1% TFA) was used. A linear gradient of mobile phase B from 10% to 95% within 27 min was used to resolve purity of the sample. Mass of the cyclic peptide was confirmed by ESI-EMM conducted by mass spectrometry facilities in the Paul Bender Chemistry Instrumentation Center (UW-Madison Department of Chemistry).

[0143] Synthesis of MCIP-2a. MCIP-2a was synthesized following reported procedures⁴¹ with the following modifications. Briefly, the linear peptide was synthesized on Rink amide resin (CEM) using a Liberty Blue HT24 System and cleaved from the resin as described above. Intramolecular disulfide bridge formation was performed by dissolving crude peptide (after cleavage and lyophilization) at 1 mg/mL in aqueous 0.1 M NH₄HCO₃ solution containing 40% DMSO. The reaction was allowed to proceed for 2 hours at room temperature with agitation. The peptide was then purified by RP-HPLC and characterized by ESI-MS as described above. Purified peptide was prepared in small aliquots, lyophilized, and stored at −80° C. until use.

ThT Fluorescence Assays

[0144] Preparation of ThT stock solution. To prepare a stock solution of ThT (2-3 mM), 2-4 mg of ThT was added to 3 mL PBS (pH 6.9) in a 15 mL conical tube. The solution was sonicated for 5 min and the supernatant was filtered through a 0.22 μm syringe filter into a 1.5 mL Eppendorf tube. This was the ThT stock solution to be used in following experiments. 2 mL of a 1:200 diluted ThT solution was prepared by adding 10 μL ThT stock solution into 1990 μL PBS. The absorbance of the diluted ThT solution measured at 412 nm was used to calculate the ThT stock concentration using an extinction coefficient (ϵ) of 31600 M^{−1} cm^{−1}.

[0145] Preparation of hIAPP stocks. A 1 mg portion of amylin trifluoroacetate salt (Bachem) was purchased as lyophilized solid in a vial. The solid was dissolved in pre-chilled 2562 μL 35 mM sodium acetate (pH 5.3) to a stock concentration of 100 μM on ice. Aliquots of hIAPP were prepared by transferring 200 μL of the 100 μM hIAPP stock solution to low-protein-binding Eppendorf tubes on ice. The aliquots were then immediately snap frozen using liquid N₂, lyophilized for 24 h, and stored at −80° C. until use.

[0146] Preparation of samples for ThT assays. Using cyclic peptide stock solution (100 μM) and PBS, 400 μL aqueous solution of cyclic peptide was prepared in a low-protein-binding Eppendorf tube to achieve the desired cyclic peptide:hIAPP molar ratio (80 μM for 4:1, 40 μM for 2:1, 20 μM for 1:1, 10 μM for 0.5:1, 0 μM for 0:1). 30 μL of the solution was then loaded to wells on a black clear-bottom low-binding 96-well plate (Corning 3881) to make a triplicate for each condition. This plate loaded with aqueous solutions of cyclic peptides was then ready for adding hIAPP.

[0147] Preparation and addition of hIAPP for ThT assays. Using ThT stock solution and PBS, 1 mL ThT solution (10 μM) was prepared in a low-protein-binding Eppendorf tube and chilled on ice. 0.5 mL of the pre-chilled ThT solution was added to one aliquot of lyophilized hIAPP, which was equilibrated to room temperature before opening the tube. Pipetting up and down four times was performed to facilitate complete dissolving of hIAPP. The 0.5 mL hIAPP-containing solution was immediately transferred to the remaining 0.5 mL ThT solution on ice. Pipetting up and down four times was performed to facilitate complete mixing. The resulting 1 mL hIAPP solution, where [hIAPP]=20 μM, was immediately transferred to a 25 mL liquid reservoir. Using a multi-channel pipette, 30 μL of the hIAPP solution was quickly loaded into wells containing aqueous solutions of cyclic peptides on the 96-well plate. Final [hIAPP]=10 μM, [ThT]=5 μM.

[0148] ThT assay conditions. The plate was then quickly covered and incubated in a Tecan M Plex microplate reader under quiescent condition at 32° C. for at least 12 h. ThT fluorescence at 480 nm was measured every 5 min through the bottom of the plate using an excitation wavelength of 440 nm.

[0149] Fitting of aggregation kinetics. ThT aggregation kinetics data was fit to a sigmoidal function using GraphPad Prism and the aggregation lag time (t_{lag}) and the apparent fibril elongation rate (k_{app}) calculated as described.⁵⁵

Transmission Electron Microscopy

[0150] Transmission electron microscopy (TEM) was performed at the University of Wisconsin School of Medicine and Public Health Electron Microscopy facility. hIAPP and cyclic peptide were dissolved in 25 mM sodium phosphate buffer (pH 6.8 with 0.4% DMSO) to achieve a molar ratio of 1:4 at concentrations of hIAPP=10 μM and cyclic peptide=40 μM. The sample was incubated quiescently at 32° C. before TEM analysis. 2 μL of the sample was placed onto Formvar-coated copper grids and the excess liquid was blotted away. 2 μL of diluted Nano-W solution was then added to stain the grids and the excess liquid was blotted away. The TEM analysis was performed on Philips CM120 at 80 kV and digital images were obtained with an AMT BioSprint12 camera.

nESI-MS Analysis

[0151] hIAPP was dissolved in 200 mM ammonium acetate buffer (pH 7.4) to prepare 100 μM hIAPP stock solution. Cyclic peptide was first dissolved in DMSO to 10 mM, then diluted into the buffer to make 1 mM cyclic peptide stock solution. These stock solutions were diluted into the buffer to achieve final concentrations of [hIAPP]=16 μM and [cyclic peptide]=64 μM. The percentage of DMSO in the final solution was 0.64% for all the samples. Prepared samples were incubated quiescently for 20 min at 32° C. Native ESI-MS analysis was then performed on a SELECT SERIES Cyclic IMS Q-TOF (Waters Corp., Wilmslow, U.K.) equipped with nano-ESI interface.

[0152] All the samples were analyzed using positive ionization ESI with a capillary voltage of 1.1 kV. The following instrumental parameters were used: source temperature 100° C.; desolvation temperature 250° C.; sampling cone 30 V; backing pressure 2.5 mbar; trap collision energy 6 V; trap DC −4 V; transfer collision energy 4 V. The system was calibrated with NaI cluster ions from a 2 μg/μL 50:50

2-propanol:water solution. Data were acquired over the m/z range of 50-8000 and processed using MassLynxV4.2 (Waters Corp., Wilmslow, U.K.).

Further Optimization of SICLOPPS-Encoding SP Through Evolution on a More Stringent Selection

[0153] We subjected the final phage pool from PACE 1 (FIG. 8B) to an additional ~250 h of PACE using a more stringent AP (pTW357d5) (FIG. 10A). This AP increases selection stringency by installing an inactivating mutation in T7c, thus requiring more copies of folded T7n to be made in order to trigger a level of gill expression sufficient for phage propagation. The stringencies of APs used in this work are ranked as: pTW006ap1a (low), pTW357b (medium), and pTW357d5 (high). Sequencing of four clonal SP from this second PACE revealed that the evolved population fixed several mutations in the intein portion of the SICLOPPS construct, including a truncation of the last 21 amino acids of the Ssp DnaE N-terminal intein (FIG. 10B). Three out of these four SP exhibited markedly increased propagation activity on cells encoding the AP42 aggregation selection compared to the parent SP (FIG. 10C). To determine if the propagation of evolved SP still depended on the activity of the cyclic peptide sequence, we cloned in a negative control sequence (CEAGQLL (SEQ ID NO:39)) where we mutated three positions important for AP42 aggregation inhibition by CKVWQLL³¹ (SEQ ID NO:30) and tested the propagation activity of the resultant phage. 3 out of 4 SP clones showed higher propagation activity when encoding CKVWQLL (SEQ ID NO:30) compared to this negative control sequence, indicating that phage fitness remained sequence-

dependent (FIG. 10C). SP clone 1 provided the largest discrimination between CKVWQLL (SEQ ID NO:30) and CEAGQLL (SEQ ID NO:39) and chosen for downstream experiments.

[0154] We characterized this more evolved SP through a series of experiments. First, we tested whether phage propagation was splicing-dependent using CKVWQLL SP where the first cysteine in the Ssp N-terminal intein was mutated to alanine, preventing formation of the first splicing intermediate (FIG. 7). As was seen from the SP isolated from PACE 1 and the parent SP, inactivation did not impact propagation of CKVWQLL SP evolved in PACE 2 (FIG. 10C) or its activity on a AP42-GFP folding reporter (FIG. 10D). Thus, we cannot rule out whether the SP-encoded cyclic peptide is being selected for activity as an intein-bound or spliced form, or both.

Abbreviations for Solid-Phase Peptide Synthesis Methods

- [0155]** Dbz=3,4-diaminobenzoic acid
- [0156]** DCM=dichloromethane
- [0157]** DIPEA=diisopropylethylamine
- [0158]** DMF=dimethylformamide
- [0159]** Fmoc=9-fluorenylmethyloxycarbonyl
- [0160]** HATU=1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
- [0161]** HPLC=high-performance liquid chromatography
- [0162]** MeCN=acetonitrile
- [0163]** Nbz=N-acylbenzimidazolinone
- [0164]** TFA=trifluoroacetic acid
- [0165]** TIPS=triisopropylsilane

SEQUENCES

PROMOTERS

P_{T7a} (attenuated T7 promoter)

(SEQ ID NO: 97)

taatacgtctcactataggg

P_{pro1}⁵⁷

(SEQ ID NO: 98)

aacaccacgtcgctccctatctgctgcctaggtctatgagtgggtgctggataaactttacgggcatgcataaggctcggtatctatattcaggagaccac

aacgggtttccctctacaaataattttgttt

P_{pro5}⁵⁷

(SEQ ID NO: 99)

aacaccacgtcgctccctatctgctgcctaggtctatgagtgggtgctggataaactttacgggcatgcataaggctcgtaggatatattcaggagaccac

aacgggtttccctctacaaataattttgttt

P_{psp}

(SEQ ID NO: 100)

gatgaaattcgccacttggttagtgtaattcgctaactcatcctggcatggtgctggtgattcttcaatcagatctttataaatcaaaaagataaaaaattg

gcacgcaaattgtattaacaggttcagcaggacaatcctgaacgcagaaatcaagaggacaacattacctgc

RIBOSOME BINDING SITES

sd5³⁶

(SEQ ID NO: 101)

aaaaaaggaaaaaa

sd8³⁶

(SEQ ID NO: 102)

aaggaggaaaaaaa

-continued

GENES

gll1 (recoded)

(SEQ ID NO: 103)

atgaaaaattattattcgcaattcctttagttgttcctttctattctcactccgctgaaactgttgaaagttgttagcaaaaccccatacagaaaattc
atttactaacgtctggaagacgacaaaacttttagatcggttacgctaactatgagggtgtctgtggaatgctacaggcgttgtagttgtactggtagc
aaactcagtggtacggtacatgggttcttattgggtctgtatccctgaaaatgagggtgggtgtctgagggtgggtgtctgagggtggcggttctgag
ggtggcggtactaaacctcctgagtagcgtgatacacctattccgggtatatttatatcaacctctcgacggcacttatccgctggtactgagcaaaa
ccccgctaatacctaactcttctcttgaggagtctcagcctcttaatactttcatgtttcagaataataggttccgaaataggcagggggcattaactgttt
atacgggcactgttactcaaggcactgaccccggtaaaaacttattaccagtacactcctgtatcatcaaaagccatgtatgacgcttactggaacggtaaa
ttcagagactgcgctttccattctggctttaatgaggatccattcggtttgtgaatatcaaggccaatcgctctgacctgcctcaacctcctgtcaatgctgg
cgggcggtctggtggtggttctggtggcggtctgagggtggtggtctgagggtggcggttctgagggtggcggtctgagggtggcggttccggtggtg
gctctggttccggtgattttgattatgaaaagatggcaacgctaataagggggctatgaccgaaaatgccgatgaaaacgcgtacagctctgacgctaaa
ggcaacttgattctgtcgctactgattacggtgctgctatcgatggtttcatgtgtgacgtttccggccttgctaattggtaatggtgctactggtgattt
tgctggtctaatcccaatggctcaagtcggtgacggtgataattcaccttttaataaatttccgtcaatattaccttccctccctcaatcggttg
aatgtcgcccttttcttcttggcgctggtaaaccttacgagttcagtatcgactgcgataagatcaacctgttccgcggtgtctttgcgtttcttttatat
gttgcacaccttatgtatgtattttctacgtttgctaacataactgcgtaataaggagttcttaa

gll1-neg

(SEQ ID NO: 104)

atgaaaaattattattcgcaattcctttagttgttcctttctattctcactccgctgaaactgttcatcaccatcaccatcacgtgaaactgttgaaag
ttgttttagcaaaaccccatacagaaaattcatttactaacgtctggaagacgacaaaacttttagatcggtacgctaactatgagggtgtctgtggaatg
ctacaggcgttgtagttttagtactggtgacgaaactcagtggttacggtacatgggttcttattgggtctgtatccctgaaaatgagggtggtggtctgag
ggtggcggttctgagggtggcggttctgagggtggcggtactaaacctcctgagtagcgtgatacacctattccgggtatacttatatcaacctctcga
cggcacttatccgctggtactgagcaaaaccccgctaatacctaactcttctcttgaggagtctcagcctcttaatactttcatgtttcagaataataggt
tccgaaataggcagggggcattaactgtttatcagggcactgttactcaaggcactgaccccggttaaaacttattaccagtacactcctgtatcatcaaaa
gccaatgtatgacgcttactggaacggtaaatcagagactgcgctttccattctggctttaatgaggatccattcggtttgtgaatatcaaggccaatcgtc
tgacctgcctcaacctcctgtcaatgctggcgcggtctggtggtggttctggtggcggtctgagggtgggtggtctgagggtggcggttctgagggtg
gcggtctgagggtggcggttccggtggtggtcttcccaaatggctcaagtcggtgacggtgataattcacctttaatgaataatttccgtcaatattta
ccttccctccctcaatcggtgaatgtcgcccttttcttcttggcgctggtaaaccttacgagttcagtatcgactgcgataagatcaacctgttccgcg
tgttttgcgtttcttttatatgttgcacaccttatgtatgtattttctacgtttgctaacataactgcgtaataaggagttcttaa

T7c R632S

(SEQ ID NO: 105)

atgaaagcatttatgcaagttgtcgaggctgacatgctctctaagggtctactcggtggcgaggcgtggtcttcgtggcataagggaagactctattcatgt
aggagtacgctgcatcgagatgctcattgagtcacccggaatggttagcttacaccgcaaaatgctggcgtagtaggtcaagactctgagactatcgaac
tcgcacctgaatacgctgaggctatcgcaaccgctgcagggtgcgtggctggcatctctccgatgttccaaccttgctagtccctcctaagccgtggact
ggcattactggtggtggtctattgggtcaacggtcgtcgtcctctggcgctggtgcgtactcacagt aagaaagcactgatgcgctacgaagacgtttacat
gcctgagggtgtacaaagcatttaacattgcgcaaaaccccgatggaaaatcaacaagaaagtcctagcgttcgccaacgtaatacacaagtggagcatt
gtccggtcgaggacatccctgcgattgagcgtgaagaactcccgatgaaaccggaagacatcgacatgaatcctgaggctctcaccgcgtggaaaacgtgct
gcgctgctgtgtacgcgaaggacaaggctcgcaagttctgcgctatcagccttgagttcatgcttgagcaagccaataagtttgcataaccataaaggccat
ctggttcccttacaacatggactggcgcggtcgtgttacctgtgtcaatgttcaaccgcaaggtaacgatatgaccaaggactgcttacgctggcga
aaggtaaaccaatcggttaaggaaaggttactactggtgaaaatccacggtgcaaacgtgctgggtgtcgataagggttccgttccctgagcgcacaaagttc
attgaggaaaaccacgagaacatcatggcttgcgctaagttccactggagaaacacttgggtgggtgagcaagattctccgttctgcttctctgcgttctg
ctttgagtagcgtgggttacagcaccacggcctgagctataactgctcccttccgctggcggttgacgggtcttgctctggcatccagcacttctccgcga

-continued

tgctccgagatgaggtaggtggtcgcgcggttaacttgcttcctagtgaaccggttcaggacatctacgggattggtgctaagaaagtcaacgagattcta
 caagcagacgcaatcaatgggaccgataaacgaagtagttaccgtgaccgatgagaacactgggtgaaatctctgagaaagtcaagctgggcactaaggcact
 ggctgggtcaatggctggcttacgggtgtactcgagtggtactaagagttcagtcagcgtggcttacgggtccaaagagttcggttcctgcaacaag
 tgctggaagataaccattcagccagctattgattccggcaagggtctgatgttactcagccgaatcaggctgctggatacatggctaagctgatttgggaa
 tctgtgagcgtgacgggtgtagctcggttgaagcaatgaactggcttaagctgctgctgctaagctgctggctgctgaggtcaaagataagaagactggaga
 gattcttcgcaagcgttgcgctgtgcatgtggtaactcctgatgggttccctgtgtggcaggaatacaagaagcctattcagacgcgcttgaacctgatgt
 tcctcggtcagttccgcttacagcctaccattaacaccaacaagatagcgagattgatgcacacaaacaggagttcggtatcgctcctaactttgtacac
 agccaagacggtagccaccttcgtaagactgtagtgtgggcacacgagaagtagcgaatcgaatctttgactgattcacgactccttcggtaccattcc
 ggctgacgctgcgaacctgttcaaagcagtcgcgaaactatggttgacacatagagtcttgtgatgtactggctgatttctacgaccagttcgctgacc
 agttgcacgagttcaattggacaaaatgccagcacttcgggctaaggtaacttgaacctcctgacatcttagagtcggacttcgcttcgcgtaa

T7c R632S/Q649S

(SEQ ID NO: 106)

atgaaagcatttatgcaagttgtcgaggtgacatgctctctaagggtctactcggtggcgaggcgtgggtcttcgtggcataaggaagactctattcatgt
 aggagtacgctgcacgagatgctcattgagtcaaccggaatgggttagcttacaccgcaaaatgctggcgtagttagtcaagactctgagactatcgaaac
 tcgcacctgaatcagctgaggtctatcgcaaccgctgcaggtgctgctggctggcgtctctccgatgttccaaccttcgctagtctctcctaagcgtggact
 ggcatctactgggtgggtgctattgggctaacggctcgctcgctcctctggcgtgggtgctgactcacagtaagaaagcactgatgcgtacgaagacgtttacat
 gcctgaggtgtacaagcgattaacattgcgcacaaacacccgatggaaaatcaacaagaaagtcctagcggctgcgaacgtaatcaccaagtggagcatt
 gtccggtcgaggacatccctgcgattgagcgtgaagaactcccgatgaaccggaagacatcgacatgaatcctgaggtctcaccgcgtggaaaacgtgct
 gccgctgctgtgtacgcaaggacaaggctcgcaagctcgcgctatcagccttgagttcatgcttgagcaagccaataagtttgctaaccataaggccat
 ctggttcccttacaacatggactggcgcggtcggtgttacctgtgtcaatgttcaaccgcgaaggtaacgatatgaccaaaaggactgcttacgctggcga
 aaggtaaaccaatcggttaaggaaggttactactgggtgaaaaatccacgggtgcaaacgtgcggtgtcgataaggttccgttccctgagcgcacaaagtcc
 attgaggaaaaccacgagaacatcatggcttcgcgtaagttccactggagaacacttgggtgggtgagcaagattctccgttctgcttcttgcgttctg
 ctttgagtacgctgggttacagcaccacggcctgagctataactgctcccttcgcgtggcgtttgacgggtcttgctctggcactccagcacttctccgca
 tgctccgagatgaggtaggtggtcgcgcggttaacttgcttcctagtgaaccggttcaggacatctacgggattggtgctaagaaagtcaacgagattcta
 caagcagacgcaatcaatgggaccgataaacgaagtagttaccgtgaccgatgagaacactgggtgaaatctctgagaaagtcaagctgggcactaaggcact
 ggctgggtcaatggctggcttacgggtgtactcgagtggtactaagagttcagtcagcgtggcttacgggtccaaagagttcggttcctgcaaaagt
 tgctggaagataaccattcagccagctattgattccggcaagggtctgatgttactcagccgaatcaggctgctggatacatggctaagctgatttgggaa
 tctgtgagcgtgacgggtgtagctcggttgaagcaatgaactggcttaagctgctgctgctaagctgctggctgctgaggtcaaagataagaagactggaga
 gattcttcgcaagcgttgcgctgtgcatgtggtaactcctgatgggttccctgtgtggcaggaatacaagaagcctattcagacgcgcttgaacctgatgt
 tcctcggtcagttccgcttacagcctaccattaacaccaacaagatagcgagattgatgcacacaaacaggagttcggtatcgctcctaactttgtacac
 agccaagacggtagccaccttcgtaagactgtagtgtgggcacacgagaagtagcgaatcgaatctttgactgattcacgactccttcggtaccattcc
 ggctgacgctgcgaacctgttcaaagcagtcgcgaaactatggttgacacatagagtcttgtgatgtactggctgatttctacgaccagttcgctgacc
 agttgcacgagttcaattggacaaaatgccagcacttcgggctaaggtaacttgaacctcctgacatcttagagtcggacttcgcttcgcgtaa

A842-T7n

GGSGG linker in underline

T7n in italics

(SEQ ID NO: 107)

atggacgcagagttccgtcacgacagtggtatgaagtcaccaccagaaattggttttttcgcccagagatggtgtagcaataaaggggctatcatcgg
 tttaatggtaggtggagttgttattgcaggtggatctgggtgtaacacgatttaacatcgctaagaaacgacttctctgacatcgaactggctgctatccggt
 tcaacactctggctgaccattacgggtgagcgtttagctcgcaacagttggcccttgagcatgagtttacgagatgggtgaagcagccttcgcaagatg
 tttgagcgtcaacttaaagctggtgaggttgaggataacgctgccgcaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactg
 gtttgaggaaagtgaagctaaagcgtgaggttgaggataacgctgccgcaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactg

-continued

ctctggcttgccctaaccagtgtgacaatacaaccgttcaggctgtagcaagcgcaatcggtcgggccattgaggacgaggctcgcttcggtcgtatccgt

gaccttgaagctaagcacttcaagaaaaacgttgaggaacaactcaacaagcgctagggcacgtctacaagtaa

AB42 F19S/L34P-T7n

GGSGG linker in underline

T7n in italics

(SEQ ID NO: 108)

atggacgcagagttccgtcacgacagtggctatgaagtcaccaccagaaattggtttctttcgccgaggatggtgtagcaataaaggggctatcatcgg

*tccgatggttaggtggagttgttattgcaggtggatctggtggt**aacacgattaacatcgctaagaacgacttctctgacatcgaactggctgctatcccg*

tcaacactctggctgaccattacgttgagcgttttagctcgcaacagttggcccttgagcatgagcttacgagatgggtgaagcacgcttcgcgaagatg

tttgagcgtcaacttaaagctggtgaggttgcggataacgctgccgcaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactg

gtttgaggaagtgaagctaagcgcggaagcgccgacagccttcaggttcctgcaagaaatcaagccggaagccgtagcgtacatcaccattaagacca

ctctggcttgccctaaccagtgtgacaatacaaccgttcaggctgtagcaagcgcaatcggtcgggccattgaggacgaggctcgcttcggtcgtatccgt

gaccttgaagctaagcacttcaagaaaaacgttgaggaacaactcaacaagcgctagggcacgtctacaagtaa

hlAPP-T7n

GGSGG linker in underline

T7n in italics

(SEQ ID NO: 109)

atgaaatgcaacactgccacatgtgcaacgcagcgctggcaaatTTTTtagttcattccagcaacaactttggtgccattctctcatctaccaacgtggg

*atccaatacatatgggtggatctggtggt**aacacgattaacatcgctaagaacgacttctctgacatcgaactggctgctatcccggttcaacactctggctg*

accattacggtgagcgttttagctcgcaacagttggcccttgagcatgagcttacgagatgggtgaagcacgcttcgcgaagatggttgagcgtcaactt

aaagctggtgaggttgcggataacgctgccgcaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactggtttgaggaagtga

agctaagcgcggaagcgccgacagccttcaggttcctgcaagaaatcaagccggaagccgtagcgtacatcaccattaagaccaactctggcttgcctaa

ccagtgtgacaatacaaccgttcaggctgtagcaagcgcaatcggtcgggccattgaggacgaggctcgcttcggtcgtatccgtgaccttgaagctaag

cacttcaagaaaaacgttgaggaacaactcaacaagcgctagggcacgtctacaagtaa

Parent SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30)

intC in bold

peptide sequence in underline

intN in italics

chitin-binding domain in bold underline

(SEQ ID NO: 110)

atggttaaagttatcggtcgtcgttccctcgaggtgcaagaatatattgatattggtcttccccagaccataattttctgctagccaatggggcgatcgc

ccacaactgtaaggttgccagttgttgtgtctcagtttcggcaccgaaatTTTaaacggttgagtacggccatttgccattggcaaaattgtgagtgaag

aaattaattgttctgtgtacagtgttgatccagaaggagagtttacaccaggcgatcgcccaatggcatgacgggggagagcaggaagtattggaatat

gaattggaagatggttcagtaatccgagctacctctgaccaccgctTTTtaaccaccgattatcaactggtggcgatcgaagaaatTTTgctaggcaact

*ggacttgttgacttttagaaaaatattaagcaaaactgaagaagctcttgacaaccatcgtcttccctttccattacttgacgctgggacaattaaa**acgacaa***

atcctgggtgtatccgcttggcaggtcaacacagcttatactgcgggacaattgggtcacatataacggcaagacgtataaatggttgcagccccacacctcc

ttggcaggatgggaacctccaacgttccctgccttgtggcagcttcaatga

PACE 2 clone 1 SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30)

intC in bold

peptide sequence in underline

truncated intN in italics; new stop codon in capital letters

chitin-binding domain in bold underline

(SEQ ID NO: 111)

atggttaaagttatcggtcgtcgttccctcgaggtgcaagaatatattgatattggtcttccccagaccataattttctgctagccaatggggcgatcgc

ccacaactgtaaggttgccagttgttgtgtctcagtttcggcaccgaaatTTTaaacggttgagtacggccatttgccattggcaaaattgtgagtgaag

aaattaattgttctgtgtacagtgttgatccaaaagggaagtttacaccaggcgatcgcccaatggcatgacgggggagagcaggaagtattggaatat

gaattggaagatggttcagtaatccgagctacctctgaccgccgctTTTtaaccaccgattatcgactggtggcgatcgaagaaatTTTgcaaggcaact

-continued

*ggacttgttgacttttagaaaaaaatTAAGcaaaactgaagaagctcttgacaaccatcgcttccctttccattacttgacgctgggacaattaaa**acgaca***

aatcctgggtgatccgcttggcagggtcaacacagcttatactgcgggacaattgggtcacatataacggcaagacgtataaatgtttgcagccccacacctc

cttggcaggatgggaaccatccaacgttctctgcttggcagcttcaatga

PACE 1 clone 2 SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30)

intC in bold

peptide sequence underline

truncated intN in italics; new stop codon in capital letters

chitin-binding domain in bold underline

(SEQ ID NO: 112)

atgggttaaagttatcggtcgctgcttccctcgagggtgcaagaatatttgatattgggtcttccccaagaccataattttctgctagccaatggggcgatcgc

ccacaactgtaaggtgtggcagttgttgtgtctcagtttcggcaccgaaatTTTaaacggttgagtacggccatttgccattggcaaaattgtgagtgaag

aaattaattgttctgtgtacagtgttgatccaaaagggagagtttacacccaggcgatcgcccaatggcatgacgggggagagcaggaagtattggaatat

gaattggaagatggttcagtaatccgagctacotctgaccaccgctttttaaccaccgattatcaactggtggcgatcgaagaaatTTTgctagggcaact

*ggacttgttgacttttagaaaaaaatTAAGcaaaactgaagaagctcttgacaaccatcgcttccctttccattacttgacgctgggacaattaaa**acgaca***

aatcctgggtgatccgcttggcagggtcaacacagcttatactgcgggacaattgggtcacatataacggcaagacgtataaatgtttgcagccccacacctc

cttggcaggatgggaaccatccaacgttctctgcttggcagcttcaatga

PACE 2 clone 1 SICLOPPS precursor for cyclo-CKVWQLL (SEQ ID NO: 30)

intC in bold

peptide sequence underline

truncated intN in italics; new stop codon in capital letters

chitin-binding domain in bold underline

(SEQ ID NO: 113)

atgggttaaagttatcggtcgctgcttccctcgagggtgcaagaatatttgatattgggtcttccccaagaccataattttctgctagccaatggggcgatcgc

ccacaactgtaaggtgtggcagttgttgtgtctcagtttcggcaccgaaatTTTaaacggttgagtacggccatttgccattggcaaaattgtgagtgaag

aaattaattgttctgtgtacagtgttgatccaaaagggaaagtTTTtacacccaggcgatcgcccaatggcatgacgggggagagcaggaagtattggaatat

gaattggaagatggttcagtaatccgagctacotctgaccgcccgtttttaaccaccgattatcgactggtggcgatcgaagaaatTTTgcaaggcaact

*ggacttgttgacttttagaaaaaaatTAAGcaaaactgaagaagctcttgacaaccatcgcttccctttccattacttgacgctgggacaattaaa**acgaca***

aatcctgggtgatccgcttggcagggtcaacacagcttatactgcgggacaattgggtcacatataacggcaagacgtataaatgtttgcagccccacacctc

cttggcaggatgggaaccatccaacgttctctgcttggcagcttcaatga

REFERENCES

- [0166] 1. F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2017, 86, 27-68.
- [0167] 2. Y. S. Eisele, C. Monteiro, C. Fearn, S. E. Encalada, R. L. Wiseman, E. T. Powers and J. W. Kelly, *Nat. Rev. Drug Discov.*, 2015, 14, 759-780.
- [0168] 3. E. E. Cawood, T. K. Karamanos, A. J. Wilson and S. E. Radford, *Biophys. Chem.*, 2021, 268, 106505.
- [0169] 4. J. C. Saunders, L. M. Young, R. A. Mahood, M. P. Jackson, C. H. Revell, R. J. Foster, D. A. Smith, A. E. Ashcroft, D. J. Brockwell and S. E. Radford, *Nat. Chem. Biol.*, 2016, 12, 94-101.
- [0170] 5. Z. Serkeny, F. Rocha, A. M. Damas, S. Macedo-Ribeiro and P. M. Martins, *Chem. Asian J.*, 2019, 14, 500-508.
- [0171] 6. Y. Xu, R. Maya-Martinez, N. Guthertz, G. R. Heath, I. W. Manfield, A. L. Breeze, F. Sobott, R. Foster and S. E. Radford, *Nat. Commun.*, 2022, 13, 1040.
- [0172] 7. A. J. Doig, M. P. del Castillo-Frias, O. Berthoumieu, B. Tarus, J. Nasica-Labouze, F. Sterpone, P. H. Nguyen, N. M. Hooper, P. Faller and P. Derreumaux, *ACS Chem. Neurosci.*, 2017, 8, 1435-1437.
- [0173] 8. Q. Nie, X. Du and M. Geng, *Acta Pharmacol. Sin.*, 2011, 32, 545-551.
- [0174] 9. A. Zorzi, K. Deyle and C. Heinis, *Curr. Opin. Chem. Biol.*, 2017, 38, 24-29.
- [0175] 10. H. Zhang and S. Chen, *RSC Chem. Biol.*, 2022, 3, 18-31.
- [0176] 11. M. R. Naylor, A. T. Bockus, M.-J. Blanco and R. S. Lokey, *Curr. Opin. Chem. Biol.*, 2017, 38, 141-147.
- [0177] 12. X. Li, T. W. Craven and P. M. Levine, *J. Med. Chem.*, 2022, 65, 11913-11926.
- [0178] 13. C. Sohrabi, A. Foster and A. Tavassoli, *Nat. Rev. Chem.*, 2020, 4, 90-101.
- [0179] 14. Y. Huang, M. M. Wiedmann and H. Suga, *Chem. Rev.*, 2018, 119, 10360-10391.
- [0180] 15. C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnou and S. J. Benkovic, *Proc. Natl. Acad. Sci.*, 1999, 96, 13638-13643.
- [0181] 16. X. Yang, K. R. Lennard, C. He, M. C. Walker, A. T. Ball, C. Doigneaux, A. Tavassoli and W. A. Van Der Donk, *Nat. Chem. Biol.*, 2018, 14, 375-380.
- [0182] 17. A. M. King, D. A. Anderson, E. Glassey, T. H. Segall-Shapiro, Z. Zhang, D. L. Niquille, A. C. Embree, K. Pratt, T. L. Williams, D. B. Gordon and others, *Nat. Commun.*, 2021, 12, 6343.
- [0183] 18. J. A. Iannuzzelli and R. Fasan, *Chem. Sci.*, 2020, 11, 6202-6208.

- [0184] 19. J. A. Kritzer, S. Hamamichi, J. M. McCaffery, S. Santagata, T. A. Naumann, K. A. Caldwell, G. A. Caldwell and S. Lindquist, *Nat. Chem. Biol.*, 2009, 5, 655-663.
- [0185] 20. D. C. Delivoria, S. Chia, J. Habchi, M. Perni, I. Matis, N. Papaevgeniou, M. Reczko, N. Chondrogianni, C. M. Dobson, M. Vendruscolo and G. Skretas, Bacterial production and direct functional screening of expanded molecular libraries for discovering inhibitors of protein aggregation, 2019, vol. 5.
- [0186] 21. I. Matis, D. C. Delivoria, B. Mavroidi, N. Papaevgeniou, S. Panoutsou, S. Bellou, K. D. Papavasileiou, Z. I. Linardaki, A. V. Stavropoulou, K. Vekrellis, N. Boukos, F. N. Kolisis, E. S. Gonos, M. Margarity, M. G. Papadopoulos, S. Efthimiopoulos, M. Pelecanou, N. Chondrogianni and G. Skretas, *Nat. Biomed. Eng.*, 2017, 1, 838-852.
- [0187] 22. M. S. Packer and D. R. Liu, *Nat. Rev. Genet.*, 2015, 16, 379-394.
- [0188] 23. K. M. Esvelt, J. C. Carlson and D. R. Liu, *Nature*, 2011, 472, 499-503.
- [0189] 24. T. B. Roth, B. M. Woolston, G. Stephanopoulos and D. R. Liu, *ACS Synth. Biol.*, 2019, 8, 796-806.
- [0190] 25. A. K. Brödel, A. Jaramillo and M. Isalan, *Nat. Commun.*, 2016, 7, 13858.
- [0191] 26. J. A. Dewey, S.-A. Azizi, V. Lu and B. C. Dickinson, *ACS Synth. Biol.*, 2021, 10, 2096-2110.
- [0192] 27. S. M. Miller, T. Wang and D. R. Liu, *Nat. Protoc.*, 2020, 15, 4101-4127.
- [0193] 28. J. Zinkus-Boltz, C. DeValk and B. C. Dickinson, *ACS Chem. Biol.*, 2019, 14, 2757-2767.
- [0194] 29. H. Cheruvu, V. L. Allen-Baume, N. M. Kad and J. M. Mason, *J. Biol. Chem.*, 2015, 290, 7426-7435.
- [0195] 30. L. L. Lee, H. Ha, Y.-T. Chang and M. P. DeLisa, *Protein Sci.*, 2009, 18, 277-286.
- [0196] 31. D. C. Delivoria, S. Chia, J. Habchi, M. Perni, I. Matis, N. Papaevgeniou, M. Reczko, N. Chondrogianni, C. M. Dobson, M. Vendruscolo and others, *Sci. Adv.*, 2019, 5, eaax5108.
- [0197] 32. I. Matis, D. C. Delivoria, B. Mavroidi, N. Papaevgeniou, S. Panoutsou, S. Bellou, K. D. Papavasileiou, Z. I. Linardaki, A. V. Stavropoulou, K. Vekrellis and Others, *Nat. Biomed. Eng.*, 2017, 1, 838-852.
- [0198] 33. T. Wang, A. H. Badran, T. P. Huang and D. R. Liu, *Nat. Chem. Biol.*, 2018, 14, 972-980.
- [0199] 34. C. Wurth, N. K. Guimard and M. H. Hecht, *J. Mol. Biol.*, 2002, 319, 1279-1290.
- [0200] 35. B. P. Hubbard, A. H. Badran, J. A. Zuris, J. P. Guilinger, K. M. Davis, L. Chen, S. Q. Tsai, J. D. Sander, J. K. Joung and D. R. Liu, *Nat. Methods*, 2015, 12, 939-942.
- [0201] 36. B. W. Thuronyi, L. W. Koblan, J. M. Levy, W.-H. Yeh, C. Zheng, G. A. Newby, C. Wilson, M. Bhaumik, O. Shubina-Oleinik, J. R. Holt and others, *Nat. Biotechnol.*, 2019, 37, 1070-1079.
- [0202] 37. A. Tavassoli, Q. Lu, J. Gam, H. Pan, S. J. Benkovic and S. N. Cohen, *ACS Chem. Biol.*, 2008, 3, 757-764.
- [0203] 38. E. Miranda, I. K. Nordgren, A. L. Male, C. E. Lawrence, F. Hoakwie, F. Cuda, W. Court, K. R. Fox, P. A. Townsend, G. K. Packham, S. A. Eccles and A. Tavassoli, *J. Am. Chem. Soc.*, 2013, 135, 10418-10425.
- [0204] 39. A. H. Badran and D. R. Liu, *Nat. Commun.*, 2015, 6, 8425.
- [0205] 40. D. Milardi, E. Gazit, S. E. Radford, Y. Xu, R. U. Gallardo, A. Caffisch, G. T. Westermarck, P. Westermarck, C. La Rosa and A. Ramamoorthy, *Chem. Rev.*, 2021, 121, 1845-1893.
- [0206] 41. A. Spanopoulou, L. Heidrich, H.-R. Chen, C. Frost, D. Hrle, E. Malideli, K. Hille, A. Grammatikopoulos, J. Bernhagen, M. Zacharias and others, *Angew. Chemie*, 2018, 130, 14711-14716.
- [0207] 42. P. Krotee, S. L. Griner, M. R. Sawaya, D. Cascio, J. A. Rodriguez, D. Shi, S. Philipp, K. Murray, L. Saelices, J. Lee and others, *J. Biol. Chem.*, 2018, 293, 2888-2902.
- [0208] 43. A. R. Horswill, S. N. Savinov and S. J. Benkovic, *Proc. Natl. Acad. Sci.*, 2004, 101, 15591-15596.
- [0209] 44. J. C. Carlson, A. H. Badran, D. A. Guggiana-Nilo and D. R. Liu, *Nat. Chem. Biol.*, 2014, 10, 216-222.
- [0210] 45. L. E. Buchanan, E. B. Dunkelberger, H. Q. Tran, P.-N. Cheng, C.-C. Chiu, P. Cao, D. P. Raleigh, J. J. De Pablo, J. S. Nowick and M. T. Zanni, *Proc. Natl. Acad. Sci.*, 2013, 110, 19285-19290.
- [0211] 46. K. Sivanessam, I. Shu, K. N. L. Huggins, M. Tatarek-Nossol, A. Kapurniotu and N. H. Andersen, *FEBS Lett.*, 2016, 590, 2575-2583.
- [0212] 47. Y. Mao, L. Yu, R. Yang, C. Ma, L. Qu and P. de B. Harrington, *Eur. J. Pharmacol.*, 2017, 804, 102-110.
- [0213] 48. C. K. Wang and D. J. Craik, *Pept. Sci.*, 2016, 106, 901-909.
- [0214] 49. N. H. Shah, G. P. Dann, M. Vila-Perello, Z. Liu and T. W. Muir, *J. Am. Chem. Soc.*, 2012, 134, 11338-11341.
- [0215] 50. J. E. Townend and A. Tavassoli, *ACS Chem. Biol.*, 2016, 11, 1624-1630.
- [0216] 51. J. Pu, J. Zinkus-Boltz and B. C. Dickinson, *Nat. Chem. Biol.*, 2017, 13, 432-438.
- [0217] 52. J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin and others, *Nature*, 2018, 556, 57-63.
- [0218] 53. S. M. Miller, T. Wang, P. B. Randolph, M. Arbab, M. W. Shen, T. P. Huang, Z. Matuszek, G. A. Newby, H. A. Rees and D. R. Liu, *Nat. Biotechnol.*, 2020, 38, 471-481.
- [0219] 54. S. M. Miller, T. Wang and D. R. Liu, *Nat. Protoc.*, 2020, 15, 4101-4127.
- [0220] 55. K. Gade Malmos, L. M. Blancas-Mejia, B. Weber, J. Buchner, M. Ramirez-Alvarado, H. Naiki and D. Otzen, *Amyloid*, 2017, 24, 1-16.
- [0221] 56. Ringquist, S. et al. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* 6, 1219-1229 (1992).
- [0222] 57. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* 39, 1131-1141 (2011).
- [0223] 58. Brödel AK, Jaramillo A, Isalan M. Engineering orthogonal dual transcription factors for multi-input synthetic promoters. *Nat Commun.* 2016 Dec. 16; 7:13858.
- [0224] 59. Segall-Shapiro T H, Meyer A J, Ellington A D, Sontag E D, Voigt C A. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Mol Syst Biol.* 2014 Jul. 30; 10(7):742.
- [0225] 60. Shis D L, Bennett M R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc Natl Acad Sci USA.* 2013 Mar. 26; 110(13):5028-33.

```

Sequence total quantity: 113
SEQ ID NO: 1      moltype = AA  length = 7
FEATURE          Location/Qualifiers
source           1..7
                 mol_type = protein
                 organism = synthetic construct
VARIANT         7
                 note = Residue may be deleted
VARIANT         6..7
                 note = Up to 29 "X" variables between position numbers 6
                   and 7 can be inserted
REGION          1..7
                 note = Circular
SEQUENCE: 1
DLGVFRX                                     7

SEQ ID NO: 2      moltype = AA  length = 7
FEATURE          Location/Qualifiers
source           1..7
                 mol_type = protein
                 organism = synthetic construct
VARIANT         7
                 note = Residue may be deleted
VARIANT         6..7
                 note = Up to 29 "X" variables between position numbers 6
                   and 7 can be inserted
REGION          1..7
                 note = Circular
SEQUENCE: 2
RCVFSGX                                     7

SEQ ID NO: 3      moltype = AA  length = 7
FEATURE          Location/Qualifiers
source           1..7
                 mol_type = protein
                 organism = synthetic construct
VARIANT         7
                 note = Residue may be deleted
VARIANT         6..7
                 note = Up to 29 "X" variables between position numbers 6
                   and 7 can be inserted
REGION          1..7
                 note = Circular
SEQUENCE: 3
HVVGVIX                                     7

SEQ ID NO: 4      moltype = AA  length = 7
FEATURE          Location/Qualifiers
source           1..7
                 mol_type = protein
                 organism = synthetic construct
VARIANT         7
                 note = Residue may be deleted
VARIANT         6..7
                 note = Up to 29 "X" variables between position numbers 6
                   and 7 can be inserted
REGION          1..7
                 note = Circular
SEQUENCE: 4
HVHSYLX                                     7

SEQ ID NO: 5      moltype = AA  length = 7

```

-continued

FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 5		
LNYPFHGX		7
SEQ ID NO: 6	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 6		
YILSIGX		7
SEQ ID NO: 7	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 7		
CGLYNIX		7
SEQ ID NO: 8	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 8		
CHSFPRX		7
SEQ ID NO: 9	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 9		
GIRSLGX		7
SEQ ID NO: 10	moltype = AA length = 7	
FEATURE	Location/Qualifiers	

-continued

source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 10		
ISCHYGX		7
SEQ ID NO: 11	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 11		
IYFHHHX		7
SEQ ID NO: 12	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 12		
VSYILLX		7
SEQ ID NO: 13	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 13		
FNLVVDX		7
SEQ ID NO: 14	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 14		
FPRGSDX		7
SEQ ID NO: 15	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	

-continued

VARIANT	mol_type = protein organism = synthetic construct 7 note = Residue may be deleted	
VARIANT	6..7 note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7 note = Circular	
SEQUENCE: 15 NRLDVSX		7
SEQ ID NO: 16 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
VARIANT	7 note = Residue may be deleted	
VARIANT	6..7 note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7 note = Circular	
SEQUENCE: 16 GLGHGNX		7
SEQ ID NO: 17 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
VARIANT	7 note = Residue may be deleted	
VARIANT	6..7 note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7 note = Circular	
SEQUENCE: 17 RVWQLCX		7
SEQ ID NO: 18 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
VARIANT	7 note = Residue may be deleted	
VARIANT	6..7 note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7 note = Circular	
SEQUENCE: 18 IVWQLCX		7
SEQ ID NO: 19 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
VARIANT	7 note = Residue may be deleted	
VARIANT	6..7 note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7 note = Circular	
SEQUENCE: 19 KVVQLAX		7
SEQ ID NO: 20 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein	

-continued

VARIANT	organism = synthetic construct	
	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 20		
RVWCARX		7
SEQ ID NO: 21	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 21		
RVYQVLX		7
SEQ ID NO: 22	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 22		
QVWSAAX		7
SEQ ID NO: 23	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 23		
RVSQVLX		7
SEQ ID NO: 24	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6	
	and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 24		
KVWGGLX		7
SEQ ID NO: 25	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	

-continued

VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 25		
RVYPVLX		7
SEQ ID NO: 26	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 26		
QVWSARX		7
SEQ ID NO: 27	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 27		
QVWCARX		7
SEQ ID NO: 28	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 28		
TVWTCLX		7
SEQ ID NO: 29	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
VARIANT	7	
	note = Residue may be deleted	
VARIANT	6..7	
	note = Up to 29 "X" variables between position numbers 6 and 7 can be inserted	
REGION	1..7	
	note = Circular	
SEQUENCE: 29		
KVYTAPX		7
SEQ ID NO: 30	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	

-continued

SEQUENCE: 30	note = Circular	
CKVWQLL		7
SEQ ID NO: 31	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 31		
CVQWLKL		7
SEQ ID NO: 32	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 32		
CKVWQLL		7
SEQ ID NO: 33	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 33		
CEVWQLL		7
SEQ ID NO: 34	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 34		
CKAWQLL		7
SEQ ID NO: 35	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
REGION	1..7	
	note = Circular	
SEQUENCE: 35		
CKVGQLL		7
SEQ ID NO: 36	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 36		
CEAWQLL		7
SEQ ID NO: 37	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 37		
CEVGQLL		7
SEQ ID NO: 38	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	

-continued

REGION	organism = synthetic construct 1..7 note = Circular	
SEQUENCE: 38 CKAGQLL		7
SEQ ID NO: 39 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 39 CEAGQLL		7
SEQ ID NO: 40 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 40 CRVWQLC		7
SEQ ID NO: 41 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 41 CIVWQLC		7
SEQ ID NO: 42 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 42 CKVWQLA		7
SEQ ID NO: 43 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 43 CRVWCAR		7
SEQ ID NO: 44 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 44 CRVYQVL		7
SEQ ID NO: 45 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 45 CQVWSAA		7

-continued

SEQ ID NO: 46	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 46		
CRVSQVL		7
SEQ ID NO: 47	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 47		
CKVWGGL		7
SEQ ID NO: 48	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 48		
CRVYPVL		7
SEQ ID NO: 49	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 49		
CQVWSAR		7
SEQ ID NO: 50	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 50		
CQVWCAR		7
SEQ ID NO: 51	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 51		
CTVWTCL		7
SEQ ID NO: 52	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 52		
CKVYTAP		7
SEQ ID NO: 53	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	

-continued

SEQUENCE: 53	note = Circular	
ARVWCAR		7
SEQ ID NO: 54	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 54		
CAVWCAR		7
SEQ ID NO: 55	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 55		
CRAWCAR		7
SEQ ID NO: 56	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 56		
CRVACAR		7
SEQ ID NO: 57	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 57		
CHVVGVI		7
SEQ ID NO: 58	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 58		
CHVHSYL		7
SEQ ID NO: 59	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 59		
CLNYPHG		7
SEQ ID NO: 60	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 60		
CYILSIG		7
SEQ ID NO: 61	moltype = AA length = 7	
FEATURE	Location/Qualifiers	

-continued

source	1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 61 CCGLYNI		7
SEQ ID NO: 62 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 62 CCHSFFR		7
SEQ ID NO: 63 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 63 CGIRSLG		7
SEQ ID NO: 64 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 64 CISCHYG		7
SEQ ID NO: 65 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 65 CIYFHHH		7
SEQ ID NO: 66 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 66 CVSYILL		7
SEQ ID NO: 67 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
SEQUENCE: 67 CHVVGVI		7
SEQ ID NO: 68 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 68 CDLGVFR		7

-continued

SEQ ID NO: 69	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 69		
CRCVFSG		7
SEQ ID NO: 70	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 70		
CLNYFHG		7
SEQ ID NO: 71	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 71		
CHVHSYL		7
SEQ ID NO: 72	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 72		
CFNLVVD		7
SEQ ID NO: 73	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 73		
CFFRGSD		7
SEQ ID NO: 74	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 74		
CGIRSLG		7
SEQ ID NO: 75	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	
	note = Circular	
SEQUENCE: 75		
CIYFHHH		7
SEQ ID NO: 76	moltype = AA length = 7	
FEATURE	Location/Qualifiers	
source	1..7	
	mol_type = protein	
	organism = synthetic construct	
REGION	1..7	

-continued

SEQUENCE: 76 CNRLDVS	note = Circular	7
SEQ ID NO: 77 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 77 CGLGHGN		7
SEQ ID NO: 78 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
SEQUENCE: 78 CDLGVFR		7
SEQ ID NO: 79 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
SEQUENCE: 79 CRCVFSG		7
SEQ ID NO: 80 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 80 AHVVGVI		7
SEQ ID NO: 81 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 81 CAVVGVI		7
SEQ ID NO: 82 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 82 CHAVGVI		7
SEQ ID NO: 83 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 83 CHVAGVI		7
SEQ ID NO: 84 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7	

-continued

SEQUENCE: 84 CHVVAVI	note = Circular	7
SEQ ID NO: 85 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 85 CHVVGAI		7
SEQ ID NO: 86 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 86 CHVVGVA		7
SEQ ID NO: 87 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 87 CLDCVFG		7
SEQ ID NO: 88 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 88 CHLDIGC		7
SEQ ID NO: 89 FEATURE source	moltype = AA length = 7 Location/Qualifiers 1..7 mol_type = protein organism = synthetic construct	
REGION	1..7 note = Circular	
SEQUENCE: 89 CLSRRCG		7
SEQ ID NO: 90 FEATURE source	moltype = DNA length = 18 Location/Qualifiers 1..18 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 90 acagttgtgg gcgatcgc		18
SEQ ID NO: 91 FEATURE source	moltype = DNA length = 45 Location/Qualifiers 1..45 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 91 acaactgtnn knnknknknk nnknknktgc tcagtttcgg caccg		45
SEQ ID NO: 92 FEATURE source	moltype = DNA length = 45 Location/Qualifiers 1..45 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 92		

-continued

acaactgtnd tndtndtndt ndtndttgtc tcagtttcgg caccg	45
SEQ ID NO: 93	moltype = DNA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 93	
ggagtgc aaa gaataattga tattgg	26
SEQ ID NO: 94	moltype = DNA length = 18
FEATURE	Location/Qualifiers
source	1..18
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 94	
caatgggcg tactcaac	18
SEQ ID NO: 95	moltype = DNA length = 59
FEATURE	Location/Qualifiers
source	1..59
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 95	
acactcttc cctacacgac gctctccga tctggagtgc aaagaatatt tgatattgg	59
SEQ ID NO: 96	moltype = DNA length = 52
FEATURE	Location/Qualifiers
source	1..52
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 96	
gtgactggag ttcagacgtg tgctctccg atctcaatgg gccgtactca ac	52
SEQ ID NO: 97	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 97	
taatacgtct cactatagg	20
SEQ ID NO: 98	moltype = DNA length = 131
FEATURE	Location/Qualifiers
source	1..131
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 98	
aacaccacgt cgtccctatc tgctgcccta ggtctatgag tggttgctgg ataactttac	60
gggcatgcat aaggctcggt atctatattc agggagacca caacgggttc cctctacaaa	120
taattttgtt t	131
SEQ ID NO: 99	moltype = DNA length = 131
FEATURE	Location/Qualifiers
source	1..131
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 99	
aacaccacgt cgtccctatc tgctgcccta ggtctatgag tggttgctgg ataactttac	60
gggcatgcat aaggctcgta ggatatattc agggagacca caacgggttc cctctacaaa	120
taattttgtt t	131
SEQ ID NO: 100	moltype = DNA length = 172
FEATURE	Location/Qualifiers
source	1..172
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 100	
gatgaaattc gccacttggt agtgtaattc gctaactcat cctggcatgt tgctgttgat	60
tcttcaatca gatctttata aatcaaaaag ataaaaaatt ggcacgcaaa ttgtattaac	120
agttcagcag gacaatcctg aacgcagaaa tcaagaggac aacattacct gc	172
SEQ ID NO: 101	moltype = DNA length = 14
FEATURE	Location/Qualifiers
source	1..14
	mol_type = other DNA

-continued

```

organism = synthetic construct

SEQUENCE: 101
aaaaaaggaa aaaa                                     14

SEQ ID NO: 102      moltype = DNA length = 15
FEATURE            Location/Qualifiers
source             1..15
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 102
aaggaggaaaa aaaaaa                                15

SEQ ID NO: 103      moltype = DNA length = 1275
FEATURE            Location/Qualifiers
source             1..1275
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 103
atgaaaaaat tattattcgc aattccttta gttgttcctt tctattctca ctccgctgaa 60
actgttgaaa gttgttttagc aaaaccccat acagaaaatt catttactaa cgtctggaaa 120
gacgacaaaa cttagatcgc ttacgctaac tatgagggct gtctgtggaa tgctacaggc 180
gtttagtatt gtactgggtg cgaaaactcag tgttacggta catgggttcc tattgggctt 240
gctatccctg aaaatgaggg tgggtggctc gaggggtggg gttctgaggg tggcggttct 300
gaggggtggc gtactaaacc tctcgagtac ggtgatacac ctattccggg ctatacttat 360
atcaaccctc tcgacggcac ttatccgect ggtactgagc aaaaccccgc taatcctaata 420
ccttctcttg aggagtctca gcctcttaata actttcatgt ttcagaataa taggttccga 480
aataggcagg gggcattaac tgtttatacg ggcactgtta ctcaaggcac tgaccccggt 540
aaaacttatt accagtacac tctgtatca tcaaaagcca tggatgacgc ttactggaac 600
ggtaaatcca gagactgcgc ttccattct ggctttaatg aggatccatt cgtttgtgaa 660
tatcaaggcc aatcgtctga cctgcctcaa cctcctgtca atgctggcgg cggtctggt 720
ggtggttctg gtggcggttc tgagggtggt ggtctgagg gtggcggttc tgagggtggc 780
ggctctgagg gagcggttc cggtggtggc tctggttccg gtgatttga ttatgaaaag 840
atggcaaacg ctaataaggg ggctatgacc gaaaatgccg atgaaacgc gctacagtct 900
gacgctaaag gcaaaactga ttctgtcgt actgattacg gtgctgctat cgatgggttc 960
attggtgacg tttcgggct tgctaattgg aatgggtgcta ctggtgattt tgctggctct 1020
aattcccaaa tggctcaagt cggtgacggt gataattcac cttaaatgaa taatttccgt 1080
caatatttac cttccctccc tcaatcggtt gaatgtcgcc ctttgtctt tggcgctggt 1140
aaaccttacg agttcagtat cgaactcgat aagatcaacc tgttcgcggg tgtctttgcg 1200
ttcttttat atgttgccac ctttatgtat gtattttcta cgtttgctaa catactgcgt 1260
aataaggagt cttaa                                     1275

SEQ ID NO: 104      moltype = DNA length = 1095
FEATURE            Location/Qualifiers
source             1..1095
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 104
atgaaaaaat tattattcgc aattccttta gttgttcctt tctattctca ctccgctgaa 60
actgttcatc accatcacca tcacgctgaa actgttgaaa gttgttttagc aaaaccccat 120
acagaaaatt catttactaa cgtctggaaa gacgacaaaa cttagatcgc ttacgctaac 180
tatgagggct gtctgtggaa tgctacaggc gtttagtatt gtactgggtg cgaaaactcag 240
tgttacggta catgggttcc tattgggctt gctatccctg aaaatgaggg tgggtggctc 300
gaggggtggc gttctgaggg tggcggttct gaggggtggc gtactaaacc tctcgagtac 360
ggtgatacac ctattccggg ctatacttat atcaaccctc tcgacggcac ttatccgect 420
ggtagctgagc aaaaccccgc taatcctaata ccttctcttg aggagtctca gcctcttaata 480
actttcatgt ttcagaataa taggttccga aataggcagg gggcattaac tgtttatacg 540
ggcactgtta ctcaaggcac tgaccccggt aaaacttatt accagtacac tctgtatca 600
tcaaaagcca tgtatgacgc ttactggaac ggtaaatcca gagactgcgc ttccattct 660
ggctttaatg aggatccatt cgtttgtgaa tatcaaggcc aatcgtctga cctgcctcaa 720
cctcctgtca atgctggcgg cggtctggtt ggtggttctg gtggcggttc tgagggtggt 780
ggctctgagg gtggcggttc tgagggtggc ggctctgagg gagcggttc cggtggtggc 840
tcttcccaaa tggctcaagt cggtgacggt gataattcac cttaaatgaa taatttccgt 900
caatatttac cttccctccc tcaatcggtt gaatgtcgcc ctttgtctt tggcgctggt 960
aaaccttacg agttcagtat cgaactcgat aagatcaacc tgttcgcggg tgtctttgcg 1020
ttcttttat atgttgccac ctttatgtat gtattttcta cgtttgctaa catactgcgt 1080
aataaggagt cttaa                                     1095

SEQ ID NO: 105      moltype = DNA length = 2118
FEATURE            Location/Qualifiers
source             1..2118
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 105
atgaaagcat ttatgcaagt tgtcgaggct gacatgctct ctaagggtct actcggtggc 60
gaggcggtgt cttcggtggc taagggaagac tctattcatg taggagtacg ctgcacgag 120
atgctcattg agtcaaccgg aatggttagc ttacaccgcc aaaaatgctg cgtagttagt 180

```

-continued

caagactctg	agactatcga	actcgcacct	gaatacgcctg	aggctatcgc	aaccctgtgca	240
ggtgcgctgg	ctggcatctc	tccgatgttc	caaccttgcg	tagttcctcc	taagccgtgg	300
actggcatta	ctgggtggtg	ctattgggct	aacggctcgc	gtcctctggc	gctgggtcgt	360
actcacagta	agaaagcact	gatgcgctac	gaagacgttt	acatgcctga	ggtgtacaaa	420
gcgattaaca	ttgcgcaaaa	caccgcatgg	aaaatcaaca	agaaagtcc	agcggtcgcc	480
aacgtaatac	ccaagtggaa	gcattgtccg	gtcaggagaca	tccctgcgat	tgagcgtgaa	540
gaactcccga	tgaaacccga	agacatcgac	atgaatcctg	aggctctcac	cgcgtggaaa	600
cgtgctgccg	ctgctgtgta	ccgcaaggac	aaggctcgca	agtctcgccg	tatcagcctt	660
gagttcatgc	ttgagcaagc	caataagttt	gctaaccata	aggccatctg	gttcccttac	720
aacatggact	ggcgcggtcg	tgtttacgct	gtgtcaatgt	tcaaccgcga	aggtaacgat	780
atgacccaaag	gactgcttac	gctggcgaaa	ggtaaaccaa	tcggtaagga	aggttactac	840
tggctgaaaa	tccacggtgc	aaactgtcgc	ggtgtcgata	aggttccgtt	ccctgagcgc	900
atcaagttca	ttgaggaaaa	ccacgagAAC	atcatggctt	gcgctaagtc	tccactggag	960
aacacttggt	gggctgagca	agattctccg	ttctgcttcc	ttgctgtctg	ctttgagtac	1020
gctggggtag	agcaccacgg	cctgagctat	aactgctccc	ttccgctggc	gtttgacggg	1080
tcttgctctg	gcattccagca	cttctccgcg	atgctccgag	atgaggtagg	tggtcgcgcg	1140
gttaacttgc	ttcctagtga	aaccgttcag	gacatctacg	ggattgttgc	taagaaagtc	1200
aacgagattc	tacaagcaga	cgaatcaaat	gggaccgata	acgaagtagt	taccgtgacc	1260
gatgagaaca	ctggtgaaat	ctctgagaaa	gtcaagctgg	gcactaaggc	actggctggt	1320
caatggctgg	cttacggtgt	tactcgcagt	gtgactaaga	gttcagtcac	gacgctggct	1380
tacgggtcca	aagagttccg	cttccgtcaa	caagtgtctg	aagataccat	tcagccagct	1440
attgattccg	gcaagggtct	gatgttcact	cagccgaatc	aggctgctgg	atacatggct	1500
aagctgattt	gggaatctgt	gagcgtgacg	gtggtagctg	cggttgaagc	aatgaactgg	1560
cttaagtctg	ctgctaagct	gctggctgct	gaggtcaaac	ataagaagac	tgagagagatt	1620
cttcgcaagc	gttgcgctgt	gcattgggta	actcctgatg	gtttccctgt	gtggcaggaa	1680
tacaagaagc	ctattcagac	gcgcttgaac	ctgatgttcc	tcggtcagtt	ccgcttacag	1740
cctaccatta	acaccaacaa	agatagcgag	attgatgcac	acaaacagga	gtctggtatc	1800
gctcctaact	ttgtacacag	ccaagacggg	agccaccttc	gtaagactgt	agtgtgggca	1860
cacgagaagt	acggaatcga	atcttttgca	ctgattcacg	actcctcgg	taccattccg	1920
gctgacgctg	cgaacctgtt	caaagcagtg	cgcgaaacta	tggttgacac	atatgagtct	1980
tgtgatgtac	tggtgatatt	ctacgaccag	ttcgcgtgac	agttgcacga	gtctcaattg	2040
gacaaaatgc	cagcacttcc	ggctaaaggt	aacttgaacc	tccgtgacat	cttagagtcg	2100
gacttcgcgt	tcgcgttaa					2118

SEQ ID NO: 106 moltype = DNA length = 2118
 FEATURE Location/Qualifiers
 source 1..2118
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 106

atgaaagcat	ttatgcaagt	tgtcagaggct	gacatgctct	ctaagggtct	actcgggtggc	60
gaggcggtgt	cttcgtggca	taaggaagac	tctattcatg	taggagtagc	ctgcacgcag	120
atgctcattg	agtcacacgg	aatgggttagc	ttacaccgcc	aaaatgctgg	cgtagtaggt	180
caagactctg	agactatcga	actcgcacct	gaatacgcctg	aggctatcgc	aaccctgtgca	240
ggtgcgctgg	ctggcatctc	tccgatgttc	caaccttgcg	tagttcctcc	taagccgtgg	300
actggcatta	ctgggtggtg	ctattgggct	aacggctcgc	gtcctctggc	gctgggtcgt	360
actcacagta	agaaagcact	gatgcgctac	gaagacgttt	acatgcctga	ggtgtacaaa	420
gcgattaaca	ttgcgcaaaa	caccgcatgg	aaaatcaaca	agaaagtcc	agcggtcgcc	480
aacgtaatac	ccaagtggaa	gcattgtccg	gtcaggagaca	tccctgcgat	tgagcgtgaa	540
gaactcccga	tgaaacccga	agacatcgac	atgaatcctg	aggctctcac	cgcgtggaaa	600
cgtgctgccg	ctgctgtgta	ccgcaaggac	aaggctcgca	agtctcgccg	tatcagcctt	660
gagttcatgc	ttgagcaagc	caataagttt	gctaaccata	aggccatctg	gttcccttac	720
aacatggact	ggcgcggtcg	tgtttacgct	gtgtcaatgt	tcaaccgcga	aggtaacgat	780
atgacccaaag	gactgcttac	gctggcgaaa	ggtaaaccaa	tcggtaagga	aggttactac	840
tggctgaaaa	tccacggtgc	aaactgtcgc	ggtgtcgata	aggttccgtt	ccctgagcgc	900
atcaagttca	ttgaggaaaa	ccacgagAAC	atcatggctt	gcgctaagtc	tccactggag	960
aacacttggt	gggctgagca	agattctccg	ttctgcttcc	ttgctgtctg	ctttgagtac	1020
gctggggtag	agcaccacgg	cctgagctat	aactgctccc	ttccgctggc	gtttgacggg	1080
tcttgctctg	gcattccagca	cttctccgcg	atgctccgag	atgaggtagg	tggtcgcgcg	1140
gttaacttgc	ttcctagtga	aaccgttcag	gacatctacg	ggattgttgc	taagaaagtc	1200
aacgagattc	tacaagcaga	cgaatcaaat	gggaccgata	acgaagtagt	taccgtgacc	1260
gatgagaaca	ctggtgaaat	ctctgagaaa	gtcaagctgg	gcactaaggc	actggctggt	1320
caatggctgg	cttacggtgt	tactcgcagt	gtgactaaga	gttcagtcac	gacgctggct	1380
tacgggtcca	aagagttccg	cttccgtcaa	agtgtgctgg	aagataccat	tcagccagct	1440
attgattccg	gcaagggtct	gatgttcact	cagccgaatc	aggctgctgg	atacatggct	1500
aagctgattt	gggaatctgt	gagcgtgacg	gtggtagctg	cggttgaagc	aatgaactgg	1560
cttaagtctg	ctgctaagct	gctggctgct	gaggtcaaac	ataagaagac	tgagagagatt	1620
cttcgcaagc	gttgcgctgt	gcattgggta	actcctgatg	gtttccctgt	gtggcaggaa	1680
tacaagaagc	ctattcagac	gcgcttgaac	ctgatgttcc	tcggtcagtt	ccgcttacag	1740
cctaccatta	acaccaacaa	agatagcgag	attgatgcac	acaaacagga	gtctggtatc	1800
gctcctaact	ttgtacacag	ccaagacggg	agccaccttc	gtaagactgt	agtgtgggca	1860
cacgagaagt	acggaatcga	atcttttgca	ctgattcacg	actcctcgg	taccattccg	1920
gctgacgctg	cgaacctgtt	caaagcagtg	cgcgaaacta	tggttgacac	atatgagtct	1980
tgtgatgtac	tggtgatatt	ctacgaccag	ttcgcgtgac	agttgcacga	gtctcaattg	2040
gacaaaatgc	cagcacttcc	ggctaaaggt	aacttgaacc	tccgtgacat	cttagagtcg	2100
gacttcgcgt	tcgcgttaa					2118

-continued

```

SEQ ID NO: 107      moltype = DNA length = 681
FEATURE            Location/Qualifiers
source              1..681
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 107
atggacgcag agttccgtca cgacagtggc tatgaagtcc accaccagaa attgggtttt 60
ttcgccgagg atgttggttag caataaagg gctatcatcg gtttaaatggt aggtggagtt 120
gttattgcag gtggatctgg tggtaacacg attaacatcg ctaagaacga cttctctgac 180
atcgaaactgg ctgctatccc gttcaacact ctggctgacc attacggtga gcgttttagct 240
cgcgaaacagt tggcccttga gcatgagtct tacgagatgg gtgaagcacg cttccgcaag 300
atgtttgagc gtcaacttaa agctgggtgag gttgcggata acgctgccgc caagcctctc 360
atcactaccc tactccctaa gatgattgca cgcatacaac actgggttga ggaagtgaag 420
gctaagcgcg gcaagcgccc gacagccttc cagttcctgc aagaaatcaa gccggaagcc 480
gtagcgtaga tcaccattaa gaccactctg gcttgccctaa ccagtgtctga caatacaacc 540
gttcaggctg tagcaagcgc aatcggtcgg gccattgagg acgaggctcg cttcggtcgt 600
atccgtgacc ttgaagctaa gcacttcaag aaaaacgttg aggaacaact caacaagcgc 660
gtagggcacg tctacaagta a                                     681

```

```

SEQ ID NO: 108      moltype = DNA length = 681
FEATURE            Location/Qualifiers
source              1..681
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 108
atggacgcag agttccgtca cgacagtggc tatgaagtcc accaccagaa attgggtttt 60
ttcgccgagg atgttggttag caataaagg gctatcatcg gtcgatggtt aggtggagtt 120
gttattgcag gtggatctgg tggtaacacg attaacatcg ctaagaacga cttctctgac 180
atcgaaactgg ctgctatccc gttcaacact ctggctgacc attacggtga gcgttttagct 240
cgcgaaacagt tggcccttga gcatgagtct tacgagatgg gtgaagcacg cttccgcaag 300
atgtttgagc gtcaacttaa agctgggtgag gttgcggata acgctgccgc caagcctctc 360
atcactaccc tactccctaa gatgattgca cgcatacaac actgggttga ggaagtgaag 420
gctaagcgcg gcaagcgccc gacagccttc cagttcctgc aagaaatcaa gccggaagcc 480
gtagcgtaga tcaccattaa gaccactctg gcttgccctaa ccagtgtctga caatacaacc 540
gttcaggctg tagcaagcgc aatcggtcgg gccattgagg acgaggctcg cttcggtcgt 600
atccgtgacc ttgaagctaa gcacttcaag aaaaacgttg aggaacaact caacaagcgc 660
gtagggcacg tctacaagta a                                     681

```

```

SEQ ID NO: 109      moltype = DNA length = 666
FEATURE            Location/Qualifiers
source              1..666
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 109
atgaaatgca acactgccac atgtgcaacg cagcgcttgg caaatttttt agttcattcc 60
agcaacaact ttggtgccaat tctctcatct accaacgtgg gatccaatac atatggtgga 120
tctggtggtta acacgattaa catcgctaag aacgacttct ctgacatcga actggctgct 180
atcccggttca acactctggc tgaccattac ggtgagcgtt tagctcgaga acagtgtggc 240
cttgagcatg agtcttaacg gatgggtgaa gcacgcttcc gcaagatggt tgagcgtcaa 300
cttaaaagctg gtgaggttgc ggataacgct gcgcaccaag ctctcatcac taccctactc 360
cctaagatga ttgacgcgat caacgactgg tttgaggaag tgaagactaa gcgcggcaag 420
cgcccgacac ccttcagatt cctgcaagaa atcaagccgg aagccgtagc gtacatcacc 480
attaagacca ctctggcttg cctaaccagt gctgacaata caaccgttca ggctgtagca 540
agcgcaatcg gtcgggcccac tgaggacgag gctcgcttcg gtcgtatccg tgaccttgaa 600
gctaagcact tcaagaaaaa cgttgaggaa caactcaaca agcgcgtagg gcacgtctac 660
aagtaa                                     666

```

```

SEQ ID NO: 110      moltype = DNA length = 657
FEATURE            Location/Qualifiers
source              1..657
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 110
atggttaaag ttatcggtcg tcgttccctc ggagtgcaaa gaatatTTga tattggtctt 60
ccccaaagacc ataattttct gctagccaat ggggcgatcg cccacaactg taaggtgtgg 120
cagttgttgt gtctcagttt cggcaccgaa attttaaccg ttgagtacgg cccattgccc 180
atttgcaaaa ttgtgagtgga agaaattaat tgttctgtgt acagtgttga tccagaaggg 240
agagttttaca cccaggcgat cgcccaatgg catgaccggg gagagcagga agtattggaa 300
tatgaattgg aagatggttc agtaatccga gctacctctg accaccgctt ttaaccacc 360
gattatcaac tgttggcgat cgaagaaatt tttgctaggc aactggactt gttgacttta 420
gaaaatatta agcaaaactga agaagctctt gacaaccatc gtcttccctt tccattactt 480
gacgctggga caattaaac gacaaatcct ggtgtatccg cttggcaggt caacacagct 540
tatactgcgg gacaattggt cacatataac ggcaagacgt ataaatgttt gcagccccc 600
acctccttgg caggatggga accatccaac gttcctgcct tgtggcagct tcaatga 657

```

-continued

```

SEQ ID NO: 111      moltype = DNA  length = 658
FEATURE
source              Location/Qualifiers
                    1..658
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 111
atggttaaag ttatcgggtcg tcgttccttc ggagtgcaaa gaattattga tattggtctt 60
ccccaagacc ataattttct gctagccaat ggggcgatcg cccacaactg taagggtgtg 120
cagttgttgt gtctcagttt cggcaccgaa attttaaccg ttgagtacgg cccattgtcc 180
attggcaaaa ttgtgagtga agaaattaat tgttctgtgt acagtgttga tccaaaaggg 240
aaagtttaca cccaggcgat cgcccaatgg catgaccggg gagagcagga agtattggaa 300
tatgaattgg aagatggttc agtaatccga gctacctctg accgccgctt tttaccacc 360
gattatcgac tgggtggcgat cgaagaaatt tttgcaaggc aactggactt gttgacttta 420
gaaaaaaatt aagcaaaactg aagaagctct tgacaacccat cgtcttccct ttccattact 480
tgacgctggg acaattaaaa cgacaaatcc tgggtgtatcc gcttggcagg tcaacacagc 540
ttatactgcg ggacaattgg tcacatataa cggcaagacg tataaatgtt tgcagcccca 600
cacctccttg gcaggatggg aaccatccaa cgttctctgccc ttgtggcagc ttcaatga 658

SEQ ID NO: 112      moltype = DNA  length = 658
FEATURE
source              Location/Qualifiers
                    1..658
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 112
atggttaaag ttatcgggtcg tcgttccttc ggagtgcaaa gaattattga tattggtctt 60
ccccaagacc ataattttct gctagccaat ggggcgatcg cccacaactg taagggtgtg 120
cagttgttgt gtctcagttt cggcaccgaa attttaaccg ttgagtacgg cccattgtcc 180
attggcaaaa ttgtgagtga agaaattaat tgttctgtgt acagtgttga tccaaaaggg 240
agagtttaca cccaggcgat cgcccaatgg catgaccggg gagagcagga agtattggaa 300
tatgaattgg aagatggttc agtaatccga gctacctctg accaccgctt tttaccacc 360
gattatcaac tgggtggcgat cgaagaaatt tttgctaggc aactggactt gttgacttta 420
gaaaaaaatt aagcaaaactg aagaagctct tgacaacccat cgtcttccct ttccattact 480
tgacgctggg acaattaaaa cgacaaatcc tgggtgtatcc gcttggcagg tcaacacagc 540
ttatactgcg ggacaattgg tcacatataa cggcaagacg tataaatgtt tgcagcccca 600
cacctccttg gcaggatggg aaccatccaa cgttctctgccc ttgtggcagc ttcaatga 658

SEQ ID NO: 113      moltype = DNA  length = 658
FEATURE
source              Location/Qualifiers
                    1..658
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 113
atggttaaag ttatcgggtcg tcgttccttc ggagtgcaaa gaattattga tattggtctt 60
ccccaagacc ataattttct gctagccaat ggggcgatcg cccacaactg taagggtgtg 120
cagttgttgt gtctcagttt cggcaccgaa attttaaccg ttgagtacgg cccattgtcc 180
attggcaaaa ttgtgagtga agaaattaat tgttctgtgt acagtgttga tccaaaaggg 240
aaagtttaca cccaggcgat cgcccaatgg catgaccggg gagagcagga agtattggaa 300
tatgaattgg aagatggttc agtaatccga gctacctctg accgccgctt tttaccacc 360
gattatcgac tgggtggcgat cgaagaaatt tttgcaaggc aactggactt gttgacttta 420
gaaaaaaatt aagcaaaactg aagaagctct tgacaacccat cgtcttccct ttccattact 480
tgacgctggg acaattaaaa cgacaaatcc tgggtgtatcc gcttggcagg tcaacacagc 540
ttatactgcg ggacaattgg tcacatataa cggcaagacg tataaatgtt tgcagcccca 600
cacctccttg gcaggatggg aaccatccaa cgttctctgccc ttgtggcagc ttcaatga 658

```

1-18. (canceled)

19. A head-to-tail cyclic peptide comprising an amino acid sequence selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HVVGVIX_n (SEQ ID NO:3), HVHSLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n (SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHHHX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLVVDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14), NRLDVXS_n (SEQ ID NO:15), GLGHGNX_n (SEQ ID NO:16), RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), KVVQLAX_n (SEQ ID NO:19), RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVVGGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26),

QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVYTAPX_n (SEQ ID NO:29) wherein X is any amino acid and n is an integer from 0-30.

20. The cyclic peptide of claim **19**, wherein the cyclic peptide has a sequence selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HVVGVIX_n (SEQ ID NO:3), HVHSLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n (SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHHHX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLVVDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14), NRLDVXS_n (SEQ ID NO:15), and GLGHGNX_n (SEQ ID NO:16).

21. The cyclic peptide of claim **19**, wherein the cyclic peptide has a sequence selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), and HVVGVIX_n (SEQ ID NO:3).

22. The cyclic peptide of claim **19**, wherein the cyclic peptide has a sequence selected from the group consisting of RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), and KVVQLAX_n (SEQ ID NO:19, RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVVWGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26), QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVVYTAPX_n (SEQ ID NO:29).

23. A method of reducing aggregation of an aggregation-prone protein, the method comprising contacting the aggregation-prone protein with a cyclic peptide as recited in claim **19**.

24. The method of claim **23**, wherein the cyclic Peptide has a sequence selected from the group consisting of DLGVFRX_n (SEQ ID NO:1), RCVFSGX_n (SEQ ID NO:2), HVVGVIX_n (SEQ ID NO:3), HVHSYLX_n (SEQ ID NO:4), LNYFHGX_n (SEQ ID NO:5), YILSIGX_n (SEQ ID NO:6), CGLYNIX_n (SEQ ID NO:7), CHSFFRX_n (SEQ ID NO:8), GIRSLGX_n (SEQ ID NO:9), ISCHYGX_n (SEQ ID NO:10), IYFHHX_n (SEQ ID NO:11), VSYILLX_n (SEQ ID NO:12), FNLVVDX_n (SEQ ID NO:13), FFRGSDX_n (SEQ ID NO:14), NRLDV SX_n (SEQ ID NO:15), and GLGHGNX_n (SEQ ID NO:16).

25. The method of claim **24**, wherein the aggregation-prone protein comprises human islet amyloid polypeptide.

26. The method of claim **25**, wherein the contacting is performed in a subject with type 2 diabetes.

27. The method of claim **23**, wherein the cyclic peptide has a sequence selected from the group consisting of RVWQLCX_n (SEQ ID NO:17), IVWQLCX_n (SEQ ID NO:18), and KVVQLAX_n (SEQ ID NO:19, RVWCARX_n (SEQ ID NO:20), RVYQVLX_n (SEQ ID NO:21), QVWSAAX_n (SEQ ID NO:22), RVSQVLX_n (SEQ ID NO:23), KVVWGLX_n (SEQ ID NO:24), RVYPVLX_n (SEQ ID NO:25), QVWSARX_n (SEQ ID NO:26), QVWCARX_n (SEQ ID NO:27), TVWTCLX_n (SEQ ID NO:28), and KVVYTAPX_n (SEQ ID NO:29).

28. The method of claim **27**, wherein the aggregation-prone protein comprises amyloid- β 42.

29. The method of claim **28**, wherein the contacting is performed in a subject with Alzheimer's disease.

30. The method of claim **23**, wherein the contacting is performed in vivo.

31. The method of claim **23**, wherein the contacting is performed in vitro.

* * * * *