



US 20230159595A1

(19) **United States**

(12) **Patent Application Publication**

Baker et al.

(10) **Pub. No.: US 2023/0159595 A1**

(43) **Pub. Date: May 25, 2023**

(54) **POLYPEPTIDES FOR USE IN SELF-ASSEMBLING PROTEIN NANOSTRUCTURES**

(71) Applicant: **University of Washington**, Seattle, WA (US)

(72) Inventors: **David Baker**, Seattle, WA (US); **Jacob B. BALE**, Seattle, WA (US); **Neil P. KING**, Seattle, WA (US); **William H. SHEFFLER**, Seattle, WA (US); **Daniel ELLIS**, Seattle, WA (US)

(21) Appl. No.: **17/934,343**

(22) Filed: **Sep. 22, 2022**

Related U.S. Application Data

(60) Continuation of application No. 16/427,493, filed on May 31, 2019, now Pat. No. 11,485,759, which is a

continuation of application No. 15/490,351, filed on Apr. 18, 2017, now Pat. No. 10,351,603, which is a division of application No. 14/930,792, filed on Nov. 3, 2015, now Pat. No. 9,630,994.

(60) Provisional application No. 62/074,167, filed on Nov. 3, 2014.

Publication Classification

(51) **Int. Cl.**
C07K 14/00 (2006.01)

(52) **U.S. Cl.**
CPC **C07K 14/00** (2013.01)

(57) **ABSTRACT**

Synthetic nanostructures, polypeptides that are useful, for example, in making synthetic nanostructures, and methods for using such synthetic nanostructures are disclosed herein.

Specification includes a Sequence Listing.

Figure 1.

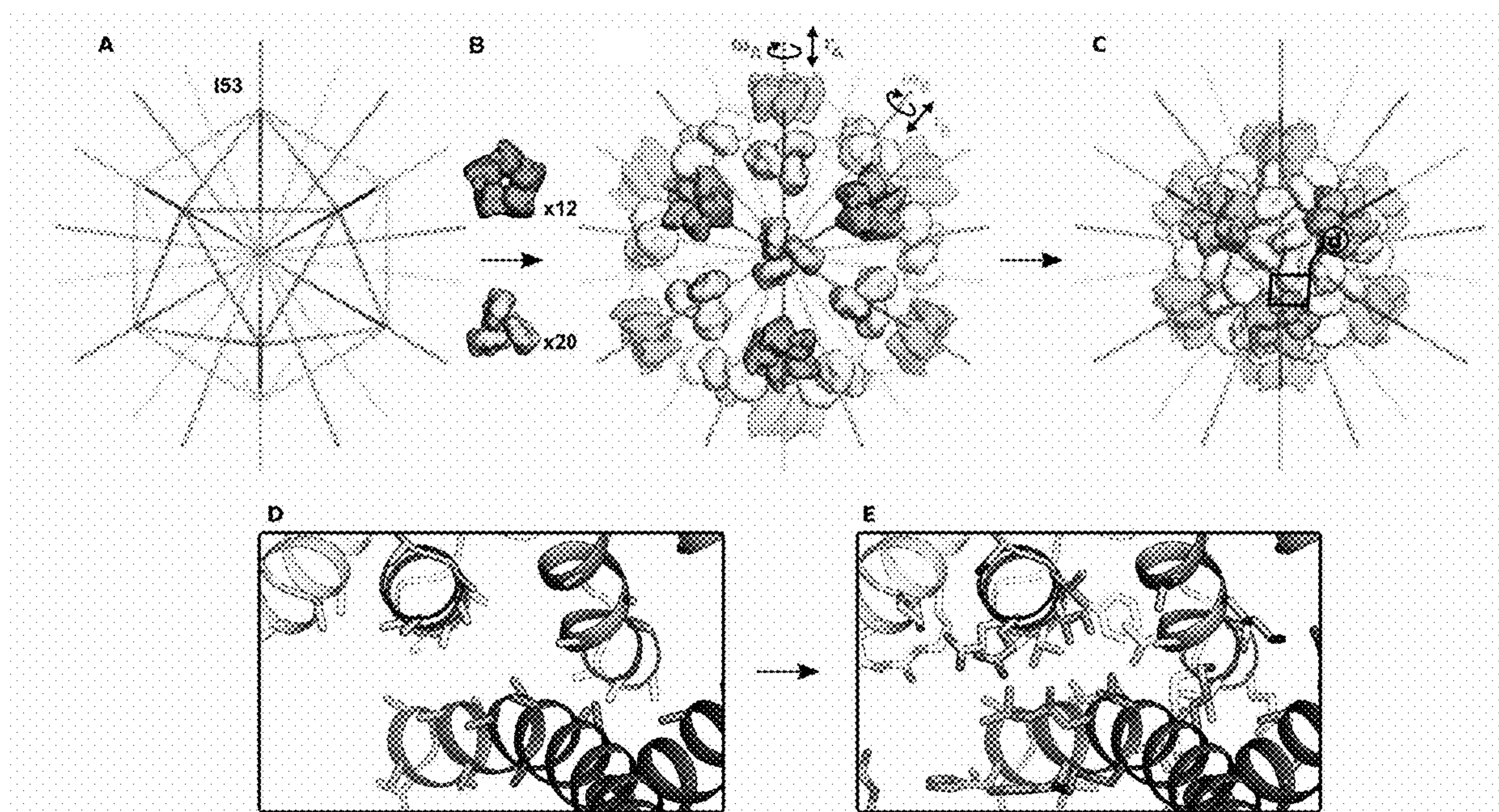


Figure 2.

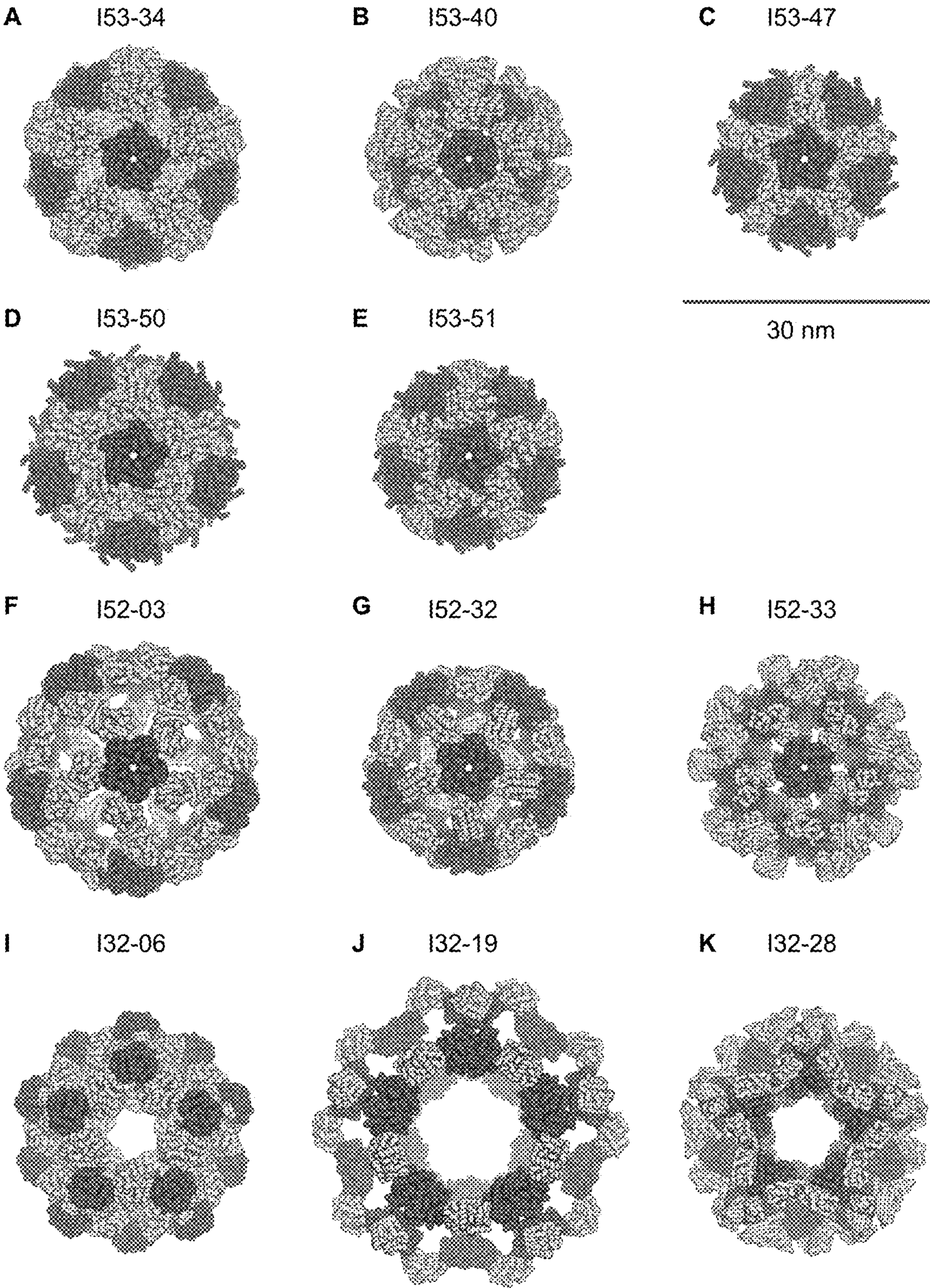
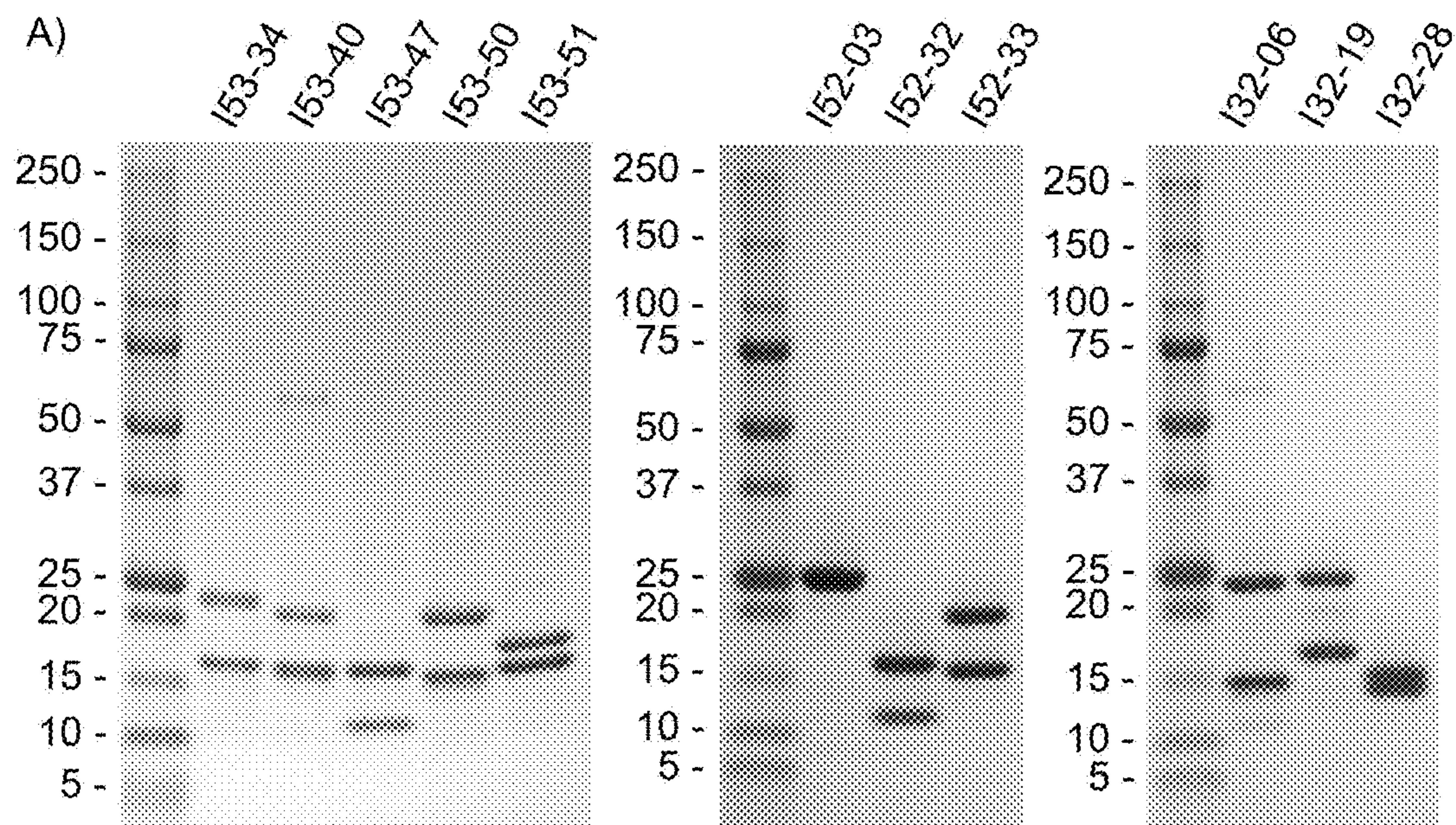


Figure 3.



B)

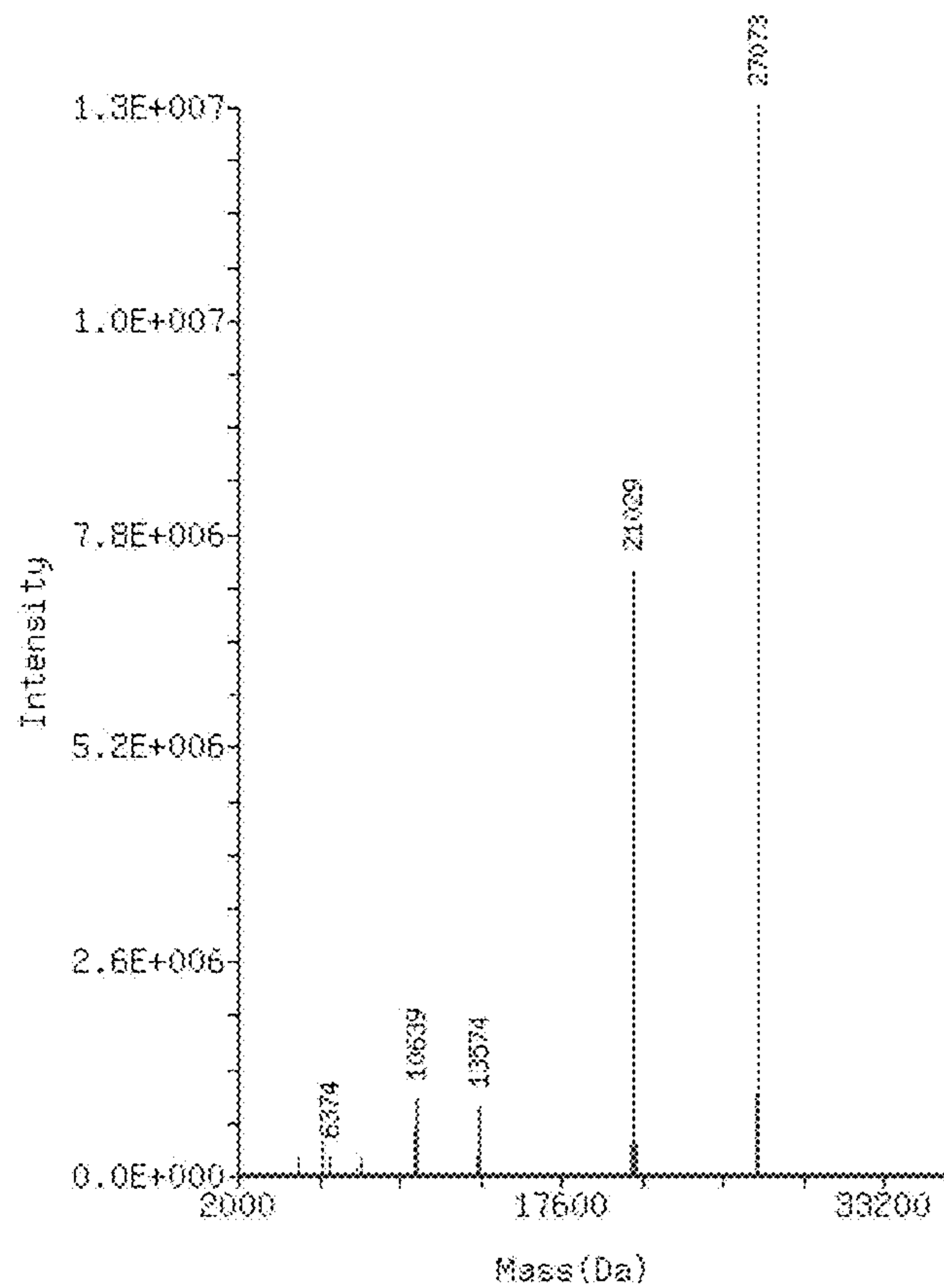


Figure 4.

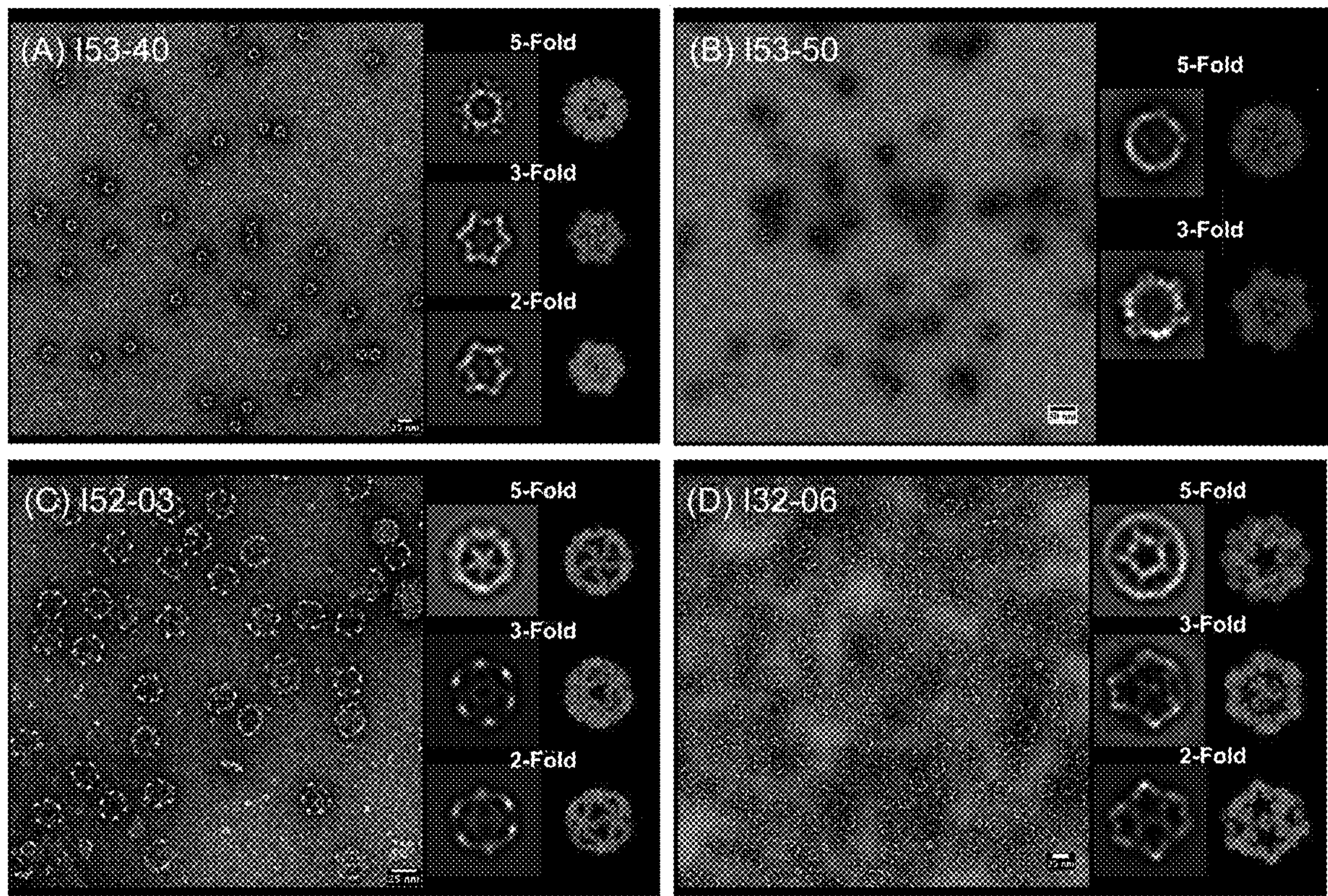


Figure 5.

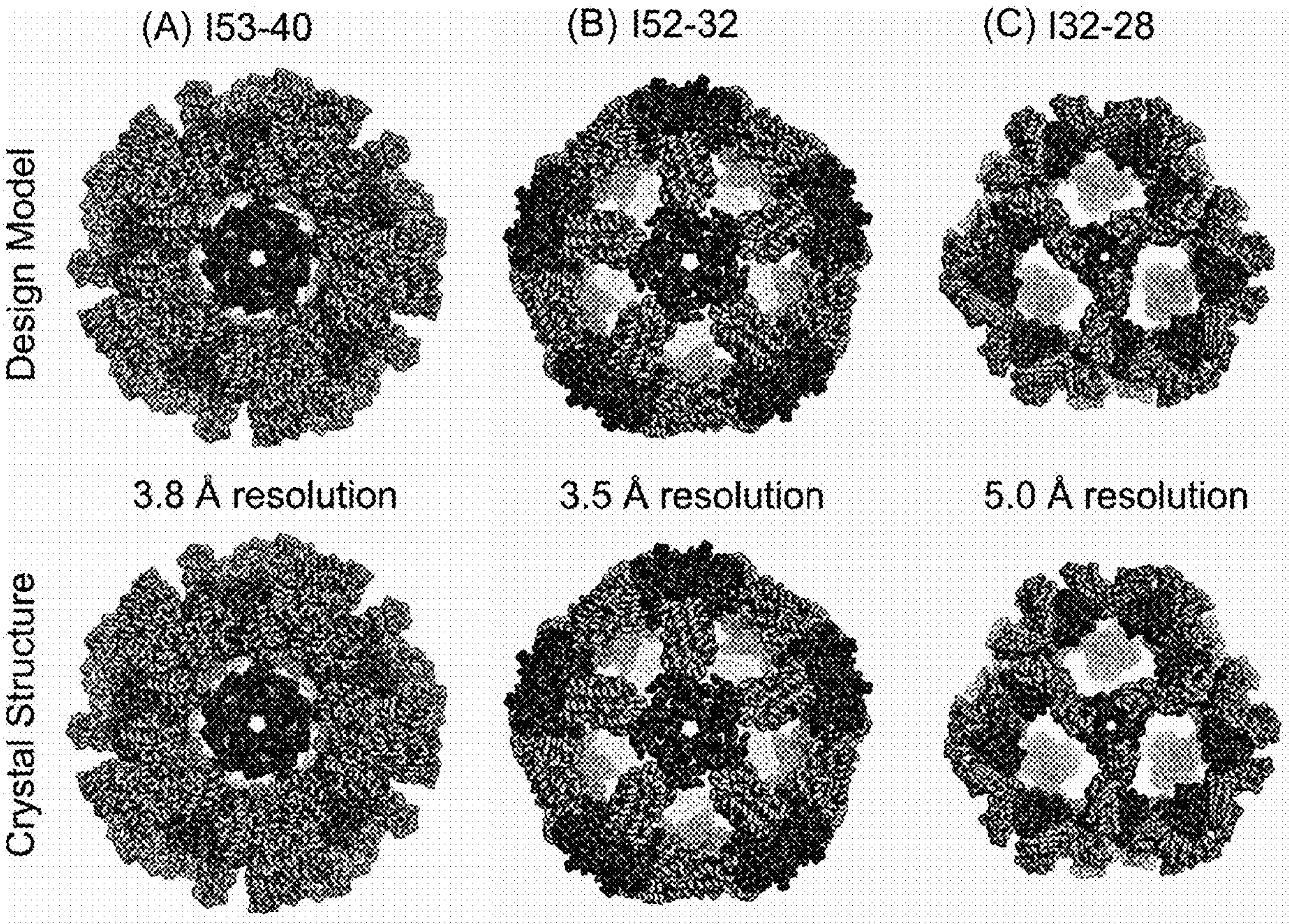


Figure 6.

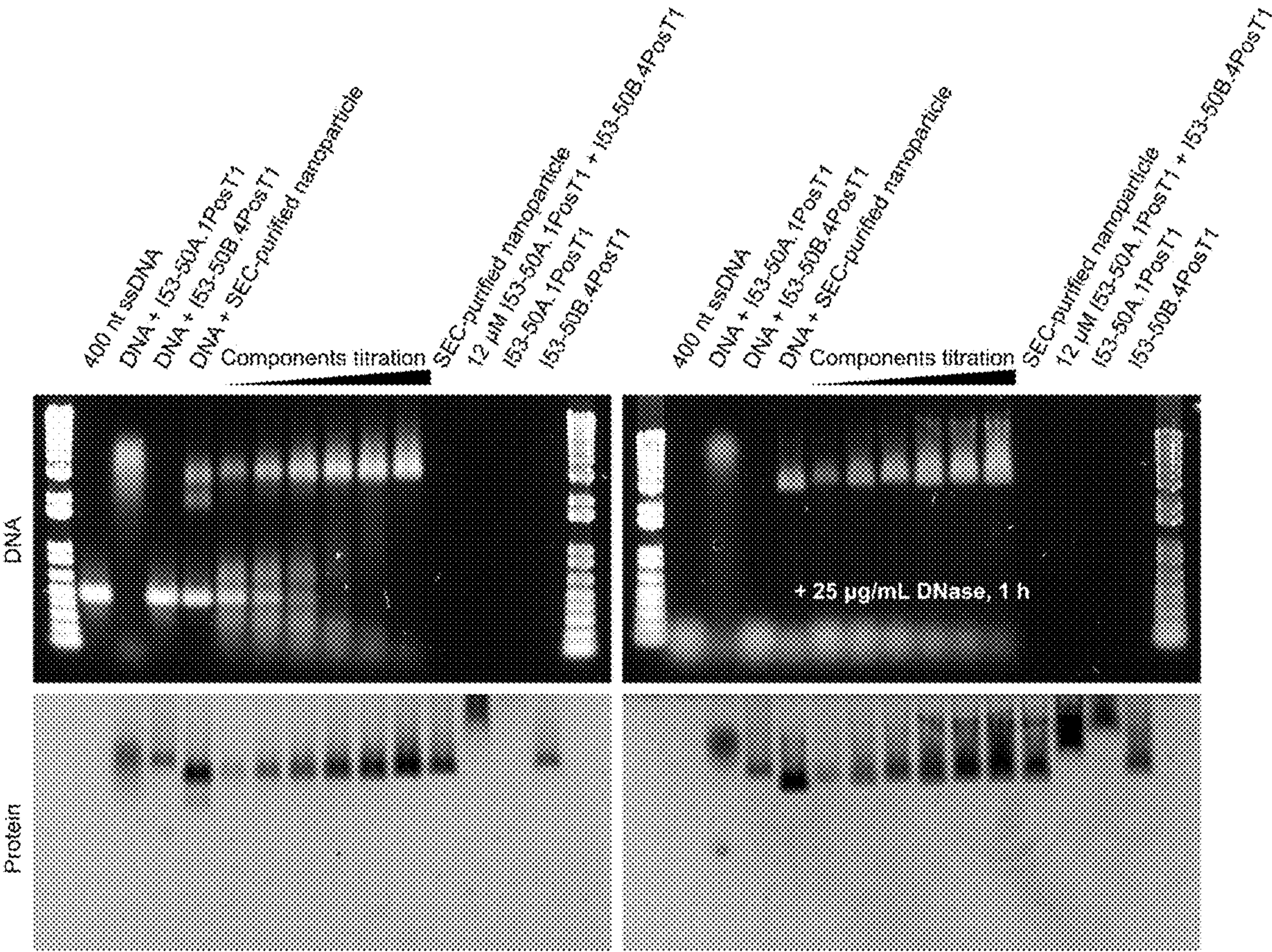
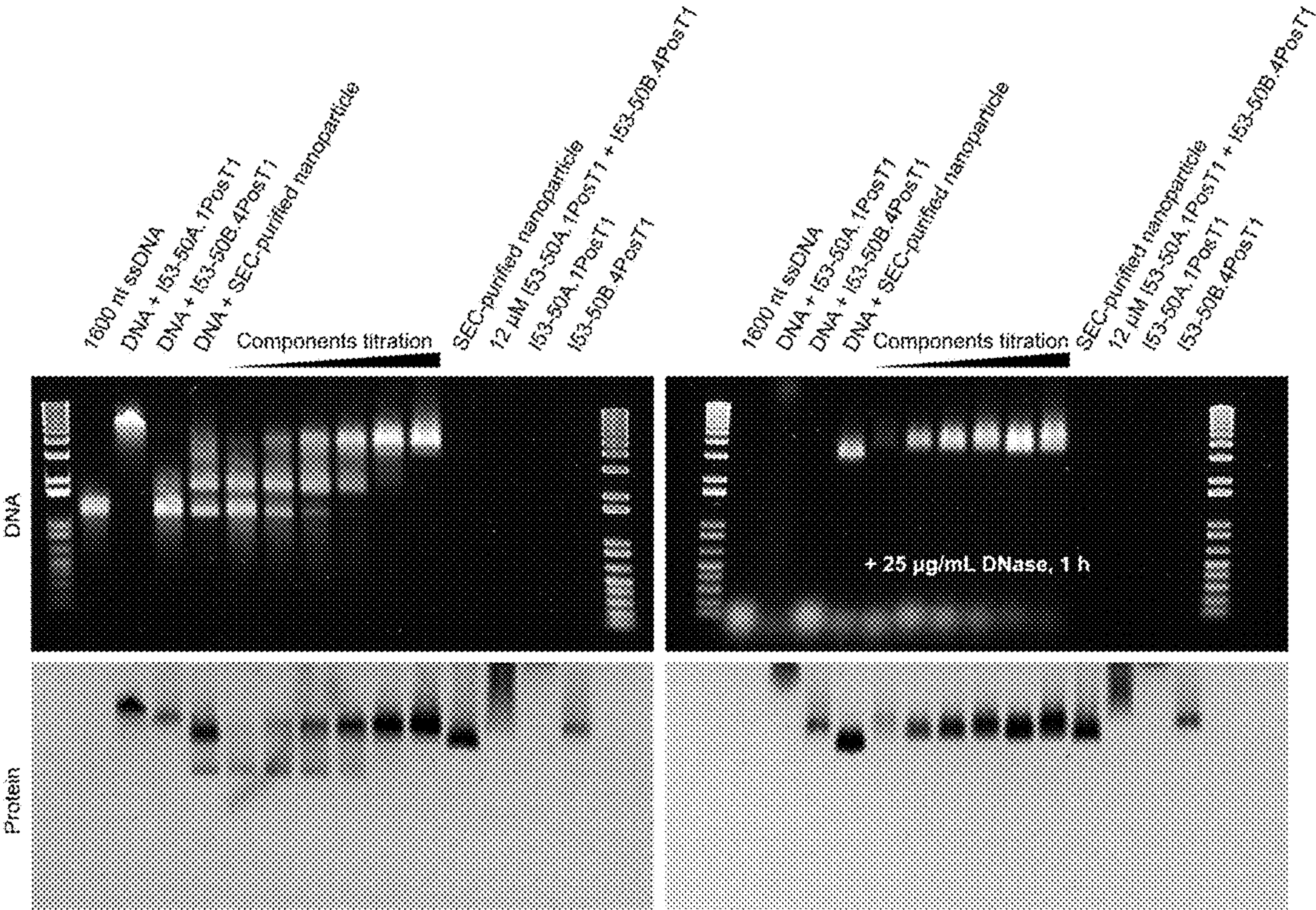


Figure 7.



POLYPEPTIDES FOR USE IN SELF-ASSEMBLING PROTEIN NANOSTRUCTURES

CROSS REFERENCE

[0001] This application claims priority to U.S. Provisional Patent Application Ser. No. 62/074,167 filed Nov. 3, 2014, incorporated by reference herein in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with U.S. government support under CHE-1332907, awarded by the National Science Foundation, and DGE-0718124, awarded by the National Science Foundation. The U.S. Government has certain rights in the invention.

BACKGROUND

[0003] Molecular self- and co-assembly of proteins into highly ordered, symmetric supramolecular complexes is an elegant and powerful means of patterning matter at the atomic scale. Recent years have seen advances in the development of self-assembling biomaterials, particularly those composed of nucleic acids. DNA has been used to create, for example, nanoscale shapes and patterns, molecular containers, and three-dimensional macroscopic crystals.

[0004] Methods for designing self-assembling proteins have progressed more slowly, yet the functional and physical properties of proteins make them attractive as building blocks for the development of advanced functional materials.

SUMMARY OF THE INVENTION

[0005] In a first aspect, the invention provides isolated polypeptides comprising an amino acid sequence that is at least 75% identical over its length, and identical at least at one identified interface position, to the amino acid sequence of a polypeptide selected from the group consisting of SEQ ID NOS:1-34.

[0006] In a second aspect, the invention provides nanostructures, comprising:

[0007] (a) a plurality of first assemblies, each first assembly comprising a plurality of identical first polypeptides, wherein the first polypeptides comprise the polypeptide of claim 1; and

[0008] (b) a plurality of second assemblies, each second assembly comprising a plurality of identical second polypeptides, wherein the second polypeptides comprise the polypeptide of claim 1, and wherein the second polypeptide differs from the first polypeptide;

[0009] wherein the plurality of first assemblies non-covalently interact with the plurality of second assemblies to form a nanostructure.

[0010] In another aspect, the present invention provides isolated nucleic acids encoding the polypeptides of the invention. In a further aspect, the invention provides nucleic acid expression vectors comprising isolated nucleic acids of the invention. In another aspect, the present invention provides recombinant host cells, comprising a nucleic acid expression vector according to the invention.

[0011] In a further aspect, the present invention provides a kit, comprising one or more isolated nanostructures of the invention; one or more of the isolated proteins of the present

invention or the assemblies of the present invention; one or more recombinant nucleic acids of the present invention; one or more recombinant expression vectors of the present invention; and/or one or more recombinant host cells of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings.

[0013] FIG. 1. Overview of the design method utilized to produce the exemplary nanostructures and sequences, illustrated with the I53 icosahedral architecture. (A) A schematic illustration of icosahedral symmetry outlined with dashed lines, with the five-fold symmetry axes shown going through each vertex and three-fold symmetry axes going through each face of the icosahedron. (B) 12 pentamers (dark grey) and 20 trimers (light grey) are aligned along the 5-fold and 3-fold symmetry axes, respectively. Each oligomer possesses two rigid body degrees of freedom, one translational (r) and one rotational (w) that are systematically sampled to identify configurations with large interfaces and high densities of contacting residues suitable for protein-protein interface design. (C) Example of such a docked configuration with a large interface and high density of contacting residues suitable for protein-protein interface design. (D) Close-up of the docked interface between the pentameric and trimeric subunits, as outlined in panel C. Side chains atoms beyond the beta carbon are ignored at this stage of design. (E) New amino acid sequences are designed at the interface to stabilize the modeled configuration.

[0014] FIG. 2. Design models of exemplary nanostructures. Computational models of the 11 exemplary nanomaterials, (A) I53-34, (B) I53-40, (C) I53-47, (D) I53-50, (E) I53-51, (F) I52-03, (G) I52-32, (H) I52-33, (I) I32-06, (J) I32-19, and (K) I32-28, are shown to scale (relative to the 30 nm scale bar), viewed down one of the 5-fold icosahedral symmetry axes with ribbon-style rendering of the protein backbone. Each I53 material comprises 12 identical pentamers (dark grey) and 20 identical trimers (light grey), each I52 material comprises 12 identical pentamers (dark grey) and 30 identical dimers (light grey), and each I32 material comprises 20 identical trimers (dark grey) and 30 identical dimers (light grey), with the designed interface formed between these oligomeric building blocks. All renderings were generated using PyMOL® Schrödinger, LLC.

[0015] FIG. 3. Sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE) and mass spectrometry analysis. Pairs of proteins encoding each material were co-expressed (as described in the Methods of Production) in *E. coli*, lysed, and purified via nickel-affinity chromatography followed by gel filtration with a Superose® 6 10/300 GL column (GE Life Sciences). (A) The resulting samples were subjected to SDS-PAGE followed by staining with GelCode® Blue Stain Reagent (Pierce Biotechnology, Inc.). The left lane in each panel contains protein molecular weight standards; the approximate molecular weights in kilodaltons are indicated directly to the left of each band. The right lanes in each panel contain the purified samples. For all of the materials except I52-03, clear bands, of similar staining intensity and near the expected molecular weights of each protein subunit, are present for each of the two

proteins comprising the purified materials. (B) While only one band (near the expected molecular weight of 27 kDa for the dimer subunit) is clearly distinguishable for I52-03 via SDS-PAGE, mass spectrometry analysis shows that the other protein subunit is also present in the sample; the mass spectrometry peak at 21,029 Da matches closely with the expected molecular weight of 21,026 Da for the pentamer subunit with loss of the initiator methionine, a common post-translational modification.

[0016] FIG. 4. Negative stain averages. Averages have been obtained of the (A) I53-40, (B) I53-50, (C) I52-03, and (D) I32-06 nanostructures and found to match well with the design models. Raw negative stain micrographs from which the averages were generated are shown on the left side of each panel. Averages (left), along with renderings from the design models (right), are shown on the right side of each panel. Views are shown corresponding approximately to the 5-fold, 3-fold, and 2-fold symmetry axes.

[0017] FIG. 5. X-ray crystallography. X-ray crystal structures (bottom) ranging from 3.5 to 5.0 Å resolution have been obtained for three of the designed materials, (A) I53-40, (B) I52-32, and (C) I32-28, and found to match closely with the design models (top). Each structure is shown using a ribbon-style rendering. Views of the I53 and I52 designs and crystal structures (panels A and B) are shown looking down one of the 5-fold symmetry axes, while the I32 design model and crystal structure (panel C) are shown looking down one of the 3-fold symmetry axes. Each crystal structure contains only a portion of the full icosahedron in the asymmetric unit. Crystal lattice symmetry was applied to generate the full icosahedra shown in the bottom panel. The I53-40 design model and crystal structure (panel A) comprise 12 pentamers (dark grey) and 20 trimers (light grey), while the I52-32 design model and crystal structure (panel B) comprise 12 pentamers (dark grey) and 30 dimers (light grey), and the I32-28 design model and crystal structure (panel C) comprise 20 trimers (dark grey) and 30 dimers (light grey). All renderings were generated using PyMOL® Schrödinger, LLC.

[0018] FIG. 6. In vitro assembly of I53-50A.1PosT1+I53-50B.4PosT1 in the presence of 400 nucleotide (nt) ssDNA leads to encapsulation and protection of the ssDNA. Mixtures of 26 ng/μL ssDNA and various proteins were analyzed by agarose gel electrophoretic mobility shift assay (EMSA) after incubation for 16 hours to determine the ability of mixtures of I53-50A.1PosT1+I53-50B.4PosT1 to encapsulate the ssDNA (left; the upper image of the gel is after staining for DNA, while the lower image of the gel is after staining for protein). Mixtures of both components (lanes labeled “Components titration” are mixtures of I53-50A.1PosT1+I53-50B.4PosT1 at 2, 4, 6, 8, 10 and 12 μM) with the DNA shift the DNA such that it migrates similarly to SEC-purified I53-50A.1PosT1+I53-50B.4PosT1 nanoparticles (upper band), while mixtures of DNA with only one protein component or the other do not. The mixtures were then incubated with 25 μg/mL DNase I for 1 hour at room temperature in order to evaluate the ability of the in vitro-assembled nanoparticles to protect the ssDNA cargo from degradation (right; the upper image of the gel is after staining for DNA, while the lower image of the gel is after staining for protein). The DNA that co-migrates with the protein in mixtures of both components (I53-50A.1PosT1+I53-50B.4PosT1; lanes labeled “Components titration” are mixtures at 2, 4, 6, 8, 10 and 12 μM) is largely protected

from DNase challenge, while free ssDNA and the mixture of ssDNA+I53-50B.4PosT1 are not. The mixture of ssDNA+I53-50A.1PosT1 is weakly protected, but migrates as a diffuse smear on the gel. Overall, the data show that the ssDNA is encapsulated in nanoparticles formed by I53-50A.1PosT1+I53-50B.4PosT1, which forms a barrier that prevents degradation of the ssDNA by DNase.

[0019] FIG. 7. In vitro assembly of I53-50A.1PosT1+I53-50B.4PosT1 in the presence of 1600 nucleotide (nt) ssDNA leads to encapsulation and protection of the ssDNA. Mixtures of 35.2 ng/μL ssDNA and various proteins were analyzed by agarose electrophoretic mobility shift assay (EMSA) after incubation for 16 hours to determine the ability of mixtures of I53-50A.1PosT1+I53-50B.4PosT1 to encapsulate the ssDNA (left; the upper image of the gel is after staining for DNA, while the lower image of the gel is after staining for protein). Mixtures of both components (lanes labeled “Components titration” are mixtures of I53-50A.1PosT1+I53-50B.4PosT1 at 2, 4, 6, 8, 10 and 12 μM) with the DNA shift the DNA such that it migrates similarly to SEC-purified I53-50A.1PosT1+I53-50B.4PosT1 nanoparticles (upper band), while mixtures of DNA with only one protein component or the other do not. The mixtures were then incubated with 25 μg/mL DNase I for 1 hour at room temperature in order to evaluate the ability of the in vitro-assembled nanoparticles to protect the ssDNA cargo from degradation (right; the upper image of the gel is after staining for DNA, while the lower image of the gel is after staining for protein). The DNA that co-migrates with the protein in mixtures of both components (I53-50A.1PosT1+I53-50B.4PosT1; lanes labeled “Components titration” are mixtures at 2, 4, 6, 8, 10 and 12 μM) is largely protected from DNase challenge, while free ssDNA and the mixture of ssDNA+I53-50B.4PosT1 are not. The mixture of ssDNA+I53-50A.1PosT1 is weakly protected, but migrates as a diffuse smear on the gel. Overall, the data show that the ssDNA is encapsulated in nanoparticles formed by I53-50A.1PosT1+I53-50B.4PosT1, which forms a barrier that prevents degradation of the ssDNA by DNase.

DETAILED DESCRIPTION

[0020] All references cited are herein incorporated by reference in their entirety. Within this application, unless otherwise stated, the techniques utilized may be found in any of several well-known references such as: *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, Calif.), “Guide to Protein Purification” in *Methods in Enzymology* (M. P. Deutscher, ed., (1990) Academic Press, Inc.); *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, Calif.), *Culture of Animal Cells: A Manual of Basic Technique*, 2nd Ed. (R. I. Freshney. 1987. Liss, Inc. New York, N.Y.), *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E. J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, Tex.).

[0021] As used herein, the singular forms “a”, “an” and “the” include plural referents unless the context clearly dictates otherwise. “And” as used herein is interchangeably used with “or” unless expressly stated otherwise.

[0022] As used herein, the amino acid residues are abbreviated as follows: alanine (Ala; A), asparagine (Asn; N),

aspartic acid (Asp; D), arginine (Arg; R), cysteine (Cys; C), glutamic acid (Glu; E), glutamine (Gln; Q), glycine (Gly; G), histidine (His; H), isoleucine (Ile; I), leucine (Leu; L), lysine (Lys; K), methionine (Met; M), phenylalanine (Phe; F), proline (Pro; P), serine (Ser; S), threonine (Thr; T), tryptophan (Trp; W), tyrosine (Tyr; Y), and valine (Val; V). As used herein, “about” means $\pm 5\%$ of the recited parameter.

[0023] All embodiments of any aspect of the invention can be used in combination, unless the context clearly dictates otherwise.

[0024] Unless the context clearly requires otherwise, throughout the description and the claims, the words ‘comprise’, ‘comprising’, and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in the sense of “including, but not limited to”. Words using the singular or plural number also include the plural and singular number, respectively. Additionally, the words “herein,” “above,” and “below” and words of similar import, when used in this application, shall refer to this application as a whole and not to any particular portions of the application.

[0025] The description of embodiments of the disclosure is not intended to be exhaustive or to limit the disclosure to the precise form disclosed. While the specific embodiments of, and examples for, the disclosure are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the disclosure, as those skilled in the relevant art will recognize.

[0026] In a first aspect, the invention provides isolated polypeptide comprising an amino acid sequence that is at least 75% identical over its length, and identical at least at one identified interface position, to the amino acid sequence of a polypeptide selected from the group consisting of SEQ ID NOS: 1-34. The isolated polypeptides of the invention can be used, for example, to prepare the nanostructures of the invention. As described in the examples that follow, the polypeptides of the invention were designed for their ability to self-assemble in pairs to form nanostructures, such as icosahedral nanostructures. The design involved design of suitable interface residues for each member of the polypep-

tide pair that can be assembled to form the nanostructure. The nanostructures of the invention include symmetrically repeated, non-natural, non-covalent polypeptide-polypeptide interfaces that orient a first assembly and a second assembly into a nanostructure, such as one with an icosahedral symmetry. Starting proteins were those derived from pentameric, trimeric, and dimeric crystal structures from the Protein Data Bank (PDB), along with a small number of crystal structures of de novo designed proteins not yet deposited in the PDB. Thus, each of the polypeptides of the present invention includes one or more modifications at “interface residues” compared to the starting proteins, permitting the polypeptides of the invention to, for example, form icosahedral nanostructures as described herein. Table 1 provides the amino acid sequence of exemplary polypeptides of the invention; the right hand column in Table 1 identifies the residue numbers in each exemplary polypeptide that were identified as present at the interface of resulting assembled nanostructures (i.e.: “identified interface residues”). As can be seen, the number of interface residues for the exemplary polypeptides of SEQ ID NO:1-34 range from 4-13. In various embodiments, the isolated polypeptides of the invention comprise an amino acid sequence that is at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical over its length, and identical at least at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 identified interface positions (depending on the number of interface residues for a given polypeptide), to the amino acid sequence of a polypeptide selected from the group consisting of SEQ ID NOS: 1-34. In other embodiments, the isolated polypeptides of the invention comprise an amino acid sequence that is at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical over its length, and identical at least at 20%, 25%, 33%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, or 100% of the identified interface positions, to the amino acid sequence of a polypeptide selected from the group consisting of SEQ ID NOS: 1-34. In further embodiments, the polypeptides of the invention comprise or consist of a polypeptide having the amino acid sequence of a polypeptide selected from the group consisting of SEQ ID NOS:1-40.

TABLE 1

Name	Amino Acid Sequence	Identified interface residues
I53-34A SEQ ID NO: 1	MEGMDPLAVLAESRLPLLTVRGGEDLAGLATVLELMGVGALEITL RTEKGLEALKALRKSGLLLGAGTVRSPKEAAEALEAGAAFLVSPGL LEEVAALAAQARGVPYLPGLVLTPIEVERALALGLSALKFFPAEPFQG VRVLRAYAEVFPVRLPTGGIKEEHLPHYAALPNLLAVGGSWLLQ GDLAAMKVKVKAALLSPQAPG	I53-34A: 28, 32, 36, 37, 186, 188, 191, 192, 195
I53-34B SEQ ID NO: 2	MTKKVGIVDTTTFARVDMAEAAIRTLKALSPNIKIIRKTVPGIKDLPV ACKKLLLEEKGCDIVMALGMPGKAEDKVCHEASLGLMLAQLMT NKHIIIEVFVHEDEAKDDDELALVRAIEHAANVYLLFKPEYLTR MAGKGLRQGREDAGPARE	I53-34B: 19, 20, 23, 24, 27, 109, 113, 116, 117, 120, 124, 148
I53-40A SEQ ID NO: 3	MTKKVGIVDTTTFARVDMASAAITLKMESPNIKIIRKTVPGIKDLPV ACKKLLLEEKGCDIVMALGMPGKAEDKVCHEASLGLMLAQLMT NKHIIIEVFVHEDEAKDDAELKILARRAIEHALNVYLLFKPEYLIR MAGKGLRQGFEDAGPARE	I53-40A: 20, 23, 24, 27, 28, 109, 112, 113, 116, 120, 124
I53-40B SEQ ID NO: 4	MSTINNQLKALKVIPVIAIDNAEDIIPLGKVLAEENLPAAEITFRS- SAA VKAIMLLRSAQPEMLIGAGTILNGVQALAAKEAGATFVVSPGFNPN TVRACQIIIGIDIVPGVNNPSTVEAALEMGLTTLKFFPAEASGGISMV KSLVGPYGDIRLMPITGGITPSNIDNYLAIPQVLACGGTWMVDKKLV TNGEWDEIARLTREIVEQVNP	I53-40B: 47, 51, 54, 58, 74, 102

TABLE 1-continued		
Name	Amino Acid Sequence	Identified interface residues
I53-47A SEQ ID NO: 5	MPIFTLNTNIKATDVPSDFLSLTSRLVGLILSKPGSYVAVHINTDQQL SFGGSTNPAAFGLMSIGGIEPSKNRDHSAVLFDHLNAMLGIPKNR MYIHFNVLNGDDVGWNGTTF	I53-47A: 22, 25, 29, 72, 79, 86, 87
I53-47B SEQ ID NO: 6	MNQHSHKDYETVRIAVVRARWHADIVDACVEAFEIAMAAIGGDRF AVDVFDVPGAYEIP LHARTLAETGRYGAVLGTAFFVNGGIYRHEFV ASAVIDGMMNVQLSTGVPVLSAVLTPHRYRDSA EHHRFFAAHFAV KGVEAARACIEILAAREKIAA	I53-47B: 28, 31, 35, 36, 39, 131, 132, 135, 139, 146
I53-50A SEQ ID NO: 7	MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVP DADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLD EEISQFCKEKGVFYMPGVMTPTLVKAMKLGHTILKLPGEVVGPO FVKAMKGPFPNVKFPVPTGGVNLDNVCEWFKAGVLAVGVGSALVK GTPDEVREKAKAFVEKIRGCTE	I53-50A: 25, 29, 33, 54, 57
I53-50B SEQ ID NO: 8	MNQHSHKDYETVRIAVVRARWHAEIVDACVSAFEAAMADIGGDR FAVDVFDVPGAYEIP LHARTLAETGRYGAVLGTAFFVNGGIYRHEF VASAVIDGMMNVQLSTGVPVLSAVLTPHRYRSDAHTLLFLALFA VKGMEAARACVEILAAREKIAA	I53-50B: 24, 28, 36, 124, 125, 127, 128, 129, 131, 132, 133, 135, 139
I53-51A SEQ ID NO: 9	MFTKSGDDGNTNVINKRVGKDSPLVNFLGDLDELNSFIGFAISKIPW EDMKDLERVQVELFEIGEDLSTQSSKKKIDESYVLWLLAATAIYRI ESGPVKLFVIPGGSEEASVLHVTRSVARRVERNNAVKYTKELPEINR MIIVYLNRLSSLLFAMALVANKRRNQSEKIYEIGKSW	I53-51 A: 80, 83, 86, 87, 88, 90, 91, 94, 166, 172, 176
I53-51B SEQ ID NO: 10	MNQHSHKDYETVRIAVVRARWHADIVDQCVRAFEEMADAGGDR FAVDVFDVPGAYEIP LHARTLAETGRYGAVLGTAFFVNGGIYRHEF VASAVIDGMMNVQLSTGVPVLSAVLTPHRYRSSREHHEFFREHFM VKGVEAAAACITILAAREKIAA	I53-51B: 31, 35, 36, 40, 122, 124, 128, 131, 135, 139, 143, 146, 147
I52-03A SEQ ID NO: 11	MGHTKGPTPQQHDGSALRIGIVHARWNKTIIMPLLIGTIAKLLECGV KASNIVVQSVPGSWELPIAVQRLYSASQLQTPSSGSPSL SAGDLLGSS TTDLTALPTTTASSTGPF DALIAIGVLIKGETMHFEYIADSVSHGLMR VQLDTGVPVIFGVLT VLTDDQAKARAGVIEGSHNHGEDWG LAAVE MGVRRRDWAAGKTE	I52-03A: 28, 32, 36, 39, 44, 49
I52-03B SEQ ID NO: 12	MYEVDHADVYDLFYLG RGKDYAAEASDIADLVR SRTPEASSLLDV ACGTGTHLEHFTKEFGDTAGLELSEDM LTHARKRLPDATLHQGDM RDFQLGRKFS AVVSMFSSVG YLKTVAELGA AVASFAEHLEPGGVV VVEPWVFPET FADGWVSADVVR RDGRTVARVSHSVREGNATRME VHFTVADPGKGV RHFSDVHLITLFHQREYEA AFMAAGLRVEYLEG GPSGRGLFVG VPA	I52-03B: 94, 115, 116, 206, 213
I52-32A SEQ ID NO: 13	MGMKEKFVLIITHGDFGKGLLSGA EVIIGKQENVHTVGLNLGDNIE KVAKEVMRIIIAKLAEDKEIIIVVDLFGGSPFNIALEMMKTFDVKVIT GINMPMLVELLTSINVYDTTELL ENISKIGKDGIKVI EKSS LKM	I52-32A: 47, 49, 53, 54, 57, 58, 61, 83, 87, 88
I52-32B SEQ ID NO: 14	MKYDGSKL RIGILHARWNLEIIAALVAGAIKRLQ EFGVKAENII IETV PGSFELPYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICD- STT HQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIEGKMHNHGEDW GAAAVEMATKFN	I52-32B: 19, 20, 23, 30, 40
I52-33A SEQ ID NO: 15	MAVKGLGEVDQKYDGSKLRIGILHARWNRKIILALVAGAVLRLL EF GVKAENII IETVPGSFELPYGSKLFVEKQKRLGKPLDAIIPIGV- LIKGS TMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIE GKMHNHGEDWGAAAVEMATKFN	I52-33A: 33, 41, 44, 50
I52-33B SEQ ID NO: 16	MGANWYLDNESSRLSFTSTKNADIAEVHRFLVLHGKVDPKGLAEV EVETESISTGIPLRDM LRLVLVFQVSKFPVAQINAQLDMRPINN LAP GAQLELR LPLTVSLRGKSHSYNAELLATRLDERRFQVVTLEPLVIHA QDFDMVRAFNALRLVAGLSAVSLSVPVGAVLIFTAR	I52-33B: 61, 63, 66, 67, 72, 147, 148, 154, 155
I32-06A SEQ ID NO: 17	MTDYIRDGSAIKALSFAIILA EADLRHIPQDLQRLAVRVIHACGMVD VANDLAFSEGAGKAGRNALLAGAPILCDARMVAEGITRSRLPADN RVIYTLSDPSVP ELAKKIGNTRSAAALDLWLPHIEGSIVAIGNAPTAL FRLFELLDAGAPKPALIIGMPVGFVGA AESKDELAANSRGVPYVIVR GRRGGSAMTAAAVNALASERE	I32-06A: 9, 12, 13, 14, 20, 30, 33, 34
I32-06B SEQ ID NO: 18	MITVFG LKSKLAPRREKLA EVIYSSLHLGLDIPKGKHAIRFLC LEKED FYYPFDRSDDYTVIEINLMAGRSEETKMLLIFLLFIALERKLGIRAH D VEITIKEQPAHCWGFGRGRTGDSARDLDYDIYV	I32-06B: 24, 71, 73, 76, 77, 80, 81, 84, 85, 88, 114, 118

TABLE 1-continued

Name	Amino Acid Sequence	Identified interface residues
I32-19A SEQ ID NO: 19	MGSDLQKLQRFSTCDISDGLLNVDNIPTGGYFPNLTAISPPQNSSIVG TAYTVLFAPIDDPRAVNYIDSVPNSILVLA- LEPHLQSQFHPFIKITQ AMYGGLMSTRAQYLKSNGTVVFGRI RDVDEHRTL NHPVFAYGVGS CAPKAVVKAVGTNVQLKILTS DGV TQTICPGDYIAGDNNGIVRIPVQ ETDISKLVTYIEKSIEVDRLVSEAIKNGLPAKAAQTARRMVLKDYI	I32-19A: 208, 213, 218, 222, 225, 226, 229, 23 3
I32-19B SEQ ID NO: 20	MSGMRVYLGADHAGYELKQAI IAF LKMTGHEPIDCGALRYDADDD YPAFCIAAATRTVADPGSLGIVLGSGNGEQIAANKVPGARCALAW SVQTAAALAREHNNAQLIGIGGRMHTLEEALRIVKAFVTTTPWSKAQR HQRRIDILAEYERTHEAPPVPGAPA	I32-19B: 20, 23, 24, 27, 117, 118, 122, 125
I32-28A SEQ ID NO: 21	MGDDARIAAIGDVDELNSQIGVLLAEPLPDDVRAALS AIQHDLFDL GGELCIPGHAAITEDHLLRLALWL VHYNGQLP PLEEFILPGGARGAA LAHVCRTVCRAERSIKALGASEPLNIAPAAYVNLLSDLLFVLARVL NRAAGGADVLWDRTRAH	I32-28A: 60, 61, 64, 67, 68, 71, 110, 120, 123, 124, 128
I32-28B SEQ ID NO: 22	MILSAEQSFTLRHPHGQAAALAFVREPAAALAGVQRLRGLSDGE QVWGELLVRVPLLGEVDLPFRSEIVRTPQGAELRPLTLTGERAWVA VSGQATAAEGGEMAFAFQFQAHLATPEABEGGGAFAFEVMVQAAA GVTL LLVAMALPQGLAAGLP PA	I32-28B: 35, 36, 54, 122, 129, 137, 140, 141, 144, 148
I53-40A.1 SEQ ID NO: 23	MTKKVGIVDTTFARVDMASAAI LTLKMESPNIKIIRKTVPGIKDLPV ACKKLEEEEGCDIVMALGMPGKKEKDKVCAHEASLGLMLAQLMT NKHII EVFVHEDEAKDDAELKILAAARAIEHALNVYYLLFKPEYLIR MAGKGLRQGFEDAGPARE	I53-40A: 20, 23, 24, 27, 28, 109, 112, 113, 116, 120, 124
I53-40B.1 SEQ ID NO: 24	MDDINNQLKRLKVIPVIAIDNAEDI IPLGKVLAENGLPAAEITFRSSA AVKAIMLLRSAQPEMLIGAGTILNGVQALAAKEAGADFV VSPGFNP NTVRACQII GIDIVPGVNNPSTVEQALEMGLTTLKFFPAEASGGISM VKS LVGPYGDIRLMPTGGITPDNIDNYLAIPQVLACGGTWMVDKKL VRNGEWDEIARLTREIVEQVNP	I53-40B: 47, 51, 54, 58, 74, 102
I53-47A.1 SEQ ID NO: 25	MPIFTLNTNIKADDVPSDFLSLTSRLVGLILSKPGSYVAVHINTDQQL SFGGSTNPAAF GTLMSIGGIEPDKNRDHSAVLF DHLNAMLGIPKNR MYIHFVNLNGDDVGWNGTTF	I53-47A: 22, 25, 29, 72, 79, 86, 87
I53-47A.1NegT 2 SEQ ID NO: 26	MPIFTLNTNIKADDVPSDFLSLTSRLVGLILSEPGSYVAVHINIDQQL SFGGSTNPAAF GTLMSIGGIEPDKNEDHSAVLF DHLNAMLGIPKNR MYIHFVDLDGDDVGWNGTTF	I53-47A: 22, 25, 29, 72, 79, 86, 87
I53-47B.1 SEQ ID NO: 27	MNQHSHKDHETVRIAVVRARWHADIVDACVEAFEIAMAAIGGDRF AVDVFDVPGAYEIP LHARTLAETGRYGAVLGTA FVVNGGIYRHEFV ASAVIDGMMNVQLDTGVPVLSAVLTPHRYRDSDEHHRFFAAHFAV KGVEAARACIEILNAREKIAA	I53-47B: 28, 31, 35, 36, 39, 131, 132, 135, 139, 146
I53-47B.1NegT 2 SEQ ID NO: 28	MNQHSHKDHETVRIAVVRARWHADIVDACVEAFEIAMAAIGGDRF AVDVFDVPGAYEIP LHARTLAETGRYGAVLGTA FVVDDGGIYDHEFV ASAVIDGMMNVQLDTGVPVLSAVLTPHEYEDSDEDHEFFAAHFAV KGVEAARACIEILNAREKIAA	I53-47B: 28, 31, 35, 36, 39, 131, 132, 135, 139, 146
I53-50A.1 SEQ ID NO: 29	MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVP DADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLD EEISQFCKEKGVFYMPGVMTPTELVKAMKLGHDILKLPGEVVG PQ FVKAMKGPPFNVK FVPTGGVNLDNVCEWFKAGVLAVGVGDALVK GDPDEVREKAKKFVEKIRGCTE	I53-50A: 25, 29, 33, 54, 57
I53-50A.1NegT 2 SEQ ID NO: 30	MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVP DADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLD EEISQFCKEKGVFYMPGVMTPTELVKAMKLGHDILKLPGEVVGPE FVEAMKGPPFNVK FVPTGGVDLDDVCEWFDAGVLAVGVGDALVE GDPDEVREDAKEFVVEIRGCTE	I53-50A: 25, 29, 33, 54, 57
I53-50A.1Post 1 SEQ ID NO: 31	MKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVP DADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLD EEISQFCKEKGVFYMPGVMTPTELVKAMKLGHDILKLPGEVVG PQ FVKAMKGPPFNVK FVPTGGVNLDNVCKWFKAGVLAVGVGKALV KGKPDEVREKAKKFVKKIRGCTE	I53-50A: 25, 29, 33, 54, 57
I53-50B.1 SEQ ID NO: 32	MNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRF AVDVFDVPGAYEIP LHARTLAETGRYGAVLGTA FVVNGGIYRHEFV ASAVIDGMMNVQLDTGVPVLSAVLTPHRYRSDAHTLLFLALFAV KGMEAARACVEILAAREKIAA	I53-50B: 24, 28, 36, 124, 125, 127, 128, 129, 131, 132, 133, 135, 139

TABLE 1-continued		
Name	Amino Acid Sequence	Identified interface residues
I53-50B.1NegT2	MNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRF AVDVFDVPGAYEIP LHARTLAETGRYGAVLGTA FVVDGGIYDHEFV ASAVIDGMMNVQLDTGVPVLSAVLTPHEYEDSDADTLLFLALFAV KGMEARACVEILAAREKIAA	I53-50B: 24, 28, 36, 124, 125, 127, 128, 129, 131, 132, 133, 135, 139
SEQ ID NO: 33		
I53-50B.4PosT1	MNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRF AVDVFDVPGAYEIP LHARTLAETGRYGAVLGTA FVVNGGIYRHEFV ASAVINGMMNVQLNTGVPVLSAVLTPHNYDKSKAHTLLFLALFAV KGMEARACVEILAAREKIAA	I53-50B: 24, 28, 36, 124, 125, 127, 128, 129, 131, 132, 133, 135, 139
SEQ ID NO: 34		

I53-40 A genus	(SEQ ID NO: 35)
MTKKVGIVDTTFARVDMASAAILTLKMESPNIKIIRKTVPGIK	
DLPVACKKLL EEGCDIVMALGMPGK (A/K) EKDKVCAHEASL	
GLMLAQLMTNKHII EVFVHEDEAKDDAELKILAARRAIEHALN	
VYYLLFKPEYLTRMAGKGLRQGFEDAGPARE	
I53-40 B genus	(SEQ ID NO: 36)
M (S/D) (T/D) INNQLK (A/R) LKVIPVIAIDNAEDI IPLGKV	
LAENGLPAAEITFRSSAAVKAIMLLRSAQPEMLIGAGTILNGV	
QALAAKEAGA (T/D) FVVS PGFNPNTVRACQIIGIDIVPGVNN	
PSTVE (A/Q) ALEMGLTTLKFFPAEASGGISMVKS LVGPYGDI	
RLMPTGGITP (S/D) NIDNYLAIPQVLACGGTWMVDKKLV	
(T/R) NGEWDEIARLTREIVEQVNP	
I53-47A genus	(SEQ ID NO: 37)
MPIFTLNTNIKA (T/D) DVPSDFLSLTSRLVGLILS (K/E) PG	
SYVAVHINTDQQLSFGGSTNPAAFGLMSIGGIEP (S/D) KN	
(R/E) DHSAVLFDHLNAMLGIPKNRMYIH FV (N/D) L (N/D) G	
DDVGWNGTTF	
I53-47B genus	(SEQ ID NO: 38)
MNQHSHKD (Y/H) ETVRIAVVRARWHADIVDACVEAFEIAMAA	
IGGDRFAVDVFDVPGAYEIP LHARTLAETGRYGAVLGTA FVV	
(N/D) GGIY (R/D) HEFVASAVIDGMMNVQL (S/D) TGVPVL	
SAVLTPH (R/E) Y (R/E) DS (A/D) E (H/D) H (R/E) FFAAH	
FAVKGVEAARACIEIL (A/N) AREKIAA	
I53-50A genus	(SEQ ID NO: 39)
MKMEELFKKKHIVAVLRANSVEEAIEKAVAVFAGGVHLIEIT	
FTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGA	
EFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGH	
(T/D) ILKLFPGEVVGP (Q/E) FV (K/E) AMKGPPFPNVKFVP	
TGGV (N/D) LD (N/D) VC (E/K) WF (K/D) AGVLAVGVG	

-continued
(S/K/D) ALV (K/E) G (T/D/K) PDEVRE (K/D) AK (A/E/K)
FV (E/K) (K/E) IRGCTE
I53-50B genus
(SEQ ID NO: 40)
MNQHSHKD (Y/H) ETVRIAVVRARWHAEIVDACVSAFEAAM
(A/R) DIGGDRFAVDVFDVPGAYEIP LHARTLAETGRYGAV
LGTA FVV (N/D) GGIY (R/D) HEFVASAVI (D/N) GMMNVQL
(S/D/N) TGVPVLSAVLTPH (R/E/N) Y (R/D/E) (D/K) S
(D/K) A (H/D) TLLFLALFAVKGMEARACVEILAAREKIAA

[0027] As is the case with proteins in general, the polypeptides are expected to tolerate some variation in the designed sequences without disrupting subsequent assembly into nanostructures: particularly when such variation comprises conservative amino acid substitutions. As used here, “conservative amino acid substitution” means that: hydrophobic amino acids (Ala, Cys, Gly, Pro, Met, Ser, Thr, Val, Ile, Leu) can only be substituted with other hydrophobic amino acids; hydrophobic amino acids with bulky side chains (Phe, Tyr, Trp) can only be substituted with other hydrophobic amino acids with bulky side chains; amino acids with positively charged side chains (Arg, His, Lys) can only be substituted with other amino acids with positively charged side chains; amino acids with negatively charged side chains (Asp, Glu) can only be substituted with other amino acids with negatively charged side chains; and amino acids with polar uncharged side chains (Ser, Thr, Asn, Gln) can only be substituted with other amino acids with polar uncharged side chains.

[0028] As will be apparent to those of skill in the art, the ability to widely modify surface amino acid residues without disruption of the polypeptide structure permits many types of modifications to endow the resulting self-assembled nanostructures with a variety of functions. In one non-limiting embodiment, the polypeptides of the invention can be modified to facilitate covalent linkage to a “cargo” of interest. In one non-limiting example, the polypeptides can be modified, such as by introduction of various cysteine residues at defined positions to facilitate linkage to one or more antigens of interest, such that a nanostructure of the polypeptides would provide a scaffold to provide a large number of antigens for delivery as a vaccine to generate an improved immune response. In some embodiments, some or all native cysteine residues that are present in the polypeptides but not intended to be used for conjugation may be

mutated to other amino acids to facilitate conjugation at defined positions. In another non-limiting embodiment, the polypeptides of the invention may be modified by linkage (covalent or non-covalent) with a moiety to help facilitate “endosomal escape.” For applications that involve delivering molecules of interest to a target cell, such as targeted delivery, a critical step can be escape from the endosome—a membrane-bound organelle that is the entry point of the delivery vehicle into the cell. Endosomes mature into lysosomes, which degrade their contents. Thus, if the delivery vehicle does not somehow “escape” from the endosome before it becomes a lysosome, it will be degraded and will not perform its function. There are a variety of lipids or organic polymers that disrupt the endosome and allow escape into the cytosol. Thus, in this embodiment, the polypeptides can be modified, for example, by introducing cysteine residues that will allow chemical conjugation of such a lipid or organic polymer to the monomer or resulting assembly surface. In another non-limiting example, the polypeptides can be modified, for example, by introducing cysteine residues that will allow chemical conjugation of fluorophores or other imaging agents that allow visualization of the nanostructures of the invention in vitro or in vivo.

[0029] Surface amino acid residues on the polypeptides can be mutated in order to improve the stability or solubility of the protein subunits or the assembled nanostructures. As will be known to one of skill in the art, if the polypeptide has significant sequence homology to an existing protein family, a multiple sequence alignment of other proteins from that family can be used to guide the selection of amino acid

mutations at non-conserved positions that can increase protein stability and/or solubility, a process referred to as consensus protein design (9).

[0030] Surface amino acid residues on the polypeptides can be mutated to positively charged (Arg, Lys) or negatively charged (Asp, Glu) amino acids in order to endow the protein surface with an overall positive or overall negative charge. In one non-limiting embodiment, surface amino acid residues on the polypeptides can be mutated to endow the interior surface of the self-assembling nanostructure with a high net charge. Such a nanostructure can then be used to package or encapsulate a cargo molecule with the opposite net charge due to the electrostatic interaction between the nanostructure interior surface and the cargo molecule. In one non-limiting embodiment, surface amino acid residues on the polypeptides can be mutated primarily to Arginine or Lysine residues in order to endow the interior surface of the self-assembling nanostructure with a net positive charge. Solutions containing the polypeptides can then be mixed in the presence of a nucleic acid cargo molecule such as a dsDNA, ssDNA, dsRNA, ssRNA, cDNA, miRNA, siRNA, shRNA, piRNA, or other nucleic acid in order to encapsulate the nucleic acid inside the self-assembling nanostructure. Such a nanostructure could be used, for example, to protect, deliver, or concentrate nucleic acids.

[0031] Table 2 lists surface amino acid residue numbers for each exemplary polypeptide of the invention denoted by SEQ ID NOS: 1-34. Thus, in various embodiments, 1 or more (at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more) of these surface residues may be modified in the polypeptides of the invention.

TABLE 2

Name	Amino Acid Sequence	Surface residues no tnearinterface
I53-34A SEQ ID NO: 1	MEGMDPLAVLAESRLLPLLT VRGGEDLAGLATVLELMGVG ALEITLRTEKGLEALKALRK SGLLLGAGTVRSPKEAEAAL EAGAAFLVSPGLLEEVAALA QARGVPYLPGVLTPIEVERA LALGLSALKFFPAEPFQGV VLRAYAEVFPFVRFLPTGGI KEEHLPHYAALPNLLAVGGS WLLQGDLAAVMKKVKAAKAL LSPQAPG	I53-34A: 6, 8, 9, 12, 14, 22, 25, 48, 49, 50, 52, 53, 56, 73, 74, 81, 94, 95, 101, 102, 103, 104, 119, 122, 137, 140, 143, 147, 150, 151, 153, 161, 162, 163, 164, 166, 167, 170, 172, 184, 193, 198, 199, 200, 202
I53-34B SEQ ID NO: 2	MTKKVGIVDTTFARVDMAEA AIRTLKALSPNIKIIRKTV GIKDLPVACKKLEEEGCDI VMALGMPGKAEKDKVCAHEA SLGLMLAQLMTNKHIEVFV HEDEAKDDDELIDILALVRAI EHAANVYLLFKPEYLTRMA GKGLRQGREDA GARE	I53-34B: 3, 12, 31, 33, 35, 36, 51, 54, 55, 56, 59, 69, 70, 71, 74, 93, 103, 106, 107, 108, 131, 132, 133, 134, 138, 142, 153
I53-40A SEQ ID NO: 3	MTKKVGIVDTTFARVDMASA AILTLKMESPNIKIIRKTV GIKDLPVACKKLEEEGCDI VMALGMPGKAEKDKVCAHEA SLGLMLAQLMTNKHIEVFV HEDEAKDDAELKILAARRAI EHALNVYLLFKPEYLIRMA GKGLRQGFEDAGPARE	I53-40A: 3, 4, 31, 33, 35, 36, 37, 51, 54, 55, 56, 57, 59, 69, 70, 71, 74, 93, 103, 106, 118, 127, 128, 131, 132, 133, 134, 135, 138, 139, 142, 150, 153
I53-40B SEQ ID NO: 4	MSTINNQLKALKVIPVIAID NAEDIIPLGKVLAEENGLPAA EITFRSSAAVKAIMLLRSAQ PEMLIGAGTILNGVQALAAK	I53-40B: 2, 3, 7, 9, 10, 12, 20, 21, 23, 26, 27, 30, 34, 38, 45, 60, 62, 75, 85, 94, 95, 122, 124, 126, 134, 139, 143,

TABLE 2-continued		
Name	Amino Acid Sequence	Surface residues no tnearinterface
	EAGATFVVSPGFNPNTVRAC QIIIGIDIVPGVNNPSTVEAA LEMGLTTLKFFPAEASGGIS MVKSLVGPYGDIRLMPTGGI TPSNIDNYLAIPQVLACGGT WMVDKKLVTNGEWDEIARLT REIVEQVNP	151, 153, 161, 163, 166, 167, 170, 172, 180, 184, 185, 186, 189, 190, 192, 193, 194, 195, 198, 201, 202, 205, 208, 209
I53-47A SEQ ID NO: 5	MPIFTLNTNIKATDVPSDFL SLTSRLVGLILSKPGSYVAV HINTDQQLSFGGSTNPAAF TLMSIGGIEPSKNRDHSAVL FDHLNAMLGIPKNRMYIH FVNLNGDDVGWNGTTF	I53-47A: 11, 13, 14, 17, 34, 36, 37, 45, 47, 54, 55, 56, 65, 69, 70, 71, 74, 91, 92, 93, 101, 103, 105, 109, 110, 112, 114
I53-47B SEQ ID NO: 6	MNQHSHKDYETVRIAVVRAR WHADIVDACVEAFEIAMA AIIGGDRFAVDVFDVPGAYE IPLHARTLAETGRYGAVLGTA FVNVNGGIYRHEFVASAVIDG MMNVQLSTGVPVLSAVLTPH RYRDSAETHRRFFAAHFAVK GVEAARACIEILAAREKIAA	I53-47B: 6, 7, 8, 9, 10, 11, 13, 18, 20, 21, 24, 43, 44, 51, 63, 67, 70, 85, 87, 101, 105, 122, 123, 124, 125, 126, 147, 152, 153, 154
I53-50A SEQ ID NO: 7	MKMEELFKKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLIE ITFTVPDADTVIKALS VLKEKGAIIGAGTVTSVEQCR KAVESGAEFIVSPHLDEEISQ FCKEKGVFYMPGVMTPTEL VKAMKMGPFPPNVKFVPTG GVNLDNVCEWFKAGVLAVG VGSALVKGTPDEVREKAKAF VEKIRGCTE	I53-50A: 4, 5, 6, 8, 9, 11, 17, 19, 23, 37, 46, 47, 59, 74, 77, 78, 81, 94, 95, 98, 101, 102, 103, 106, 119, 122, 126, 139, 142, 145, 149, 150, 152, 160, 161, 162, 163, 166, 169, 179, 183, 185, 188, 191, 192, 194, 198, 199
I53-50B SEQ ID NO: 8	MNQHSHKDYETVRIAVVRAR WHAIEIVDACVSAFEAAMADI GGDRFAVDVFDVPGAYE IPLHARTLAETGRYGAVLGTA FVNVNGGIYRHEFVASAVIDG MMNVQLSTGVPVLSAVLTPH RYRDSAHTLLFLALFAVK GMEAARACVEILAAREKIAA	I53-50B: 6, 7, 8, 9, 10, 11, 13, 18, 20, 21, 34, 38, 39, 40, 43, 44, 48, 51, 63, 67, 70, 87, 101, 105, 118, 143, 147, 152, 153, 154
I53-51A SEQ ID NO: 9	MFTKSGDDGNTNVINKRVGK DSPLVNFLGDLDELNSFIGF AISKIPWEDMKDLERVQVE LFEIGEDLSTQSSKKKID ESYVLWLLAATAIYRIESGP VKLFVIPGGSEEASVLHVTR SVARRVERNAVKYTKELPE INRMIIVYLNRLSSLLFAMA LVA NKRRNQSEKIYEIGKSW	I53-51A: 19, 20, 24, 28, 46, 47, 51, 70, 71, 73, 74, 75, 76, 102, 122, 130, 133, 134, 135, 136, 137, 140, 162, 163, 164, 165, 169, 175, 177
I53-51B SEQ ID NO: 10	MNQHSHKDYETVRIAVVRAR WHADIVDQCVRAFEEMADA GGDRFAVDVFDVPGAYE IPLHARTLAETGRYGAVLGTA FVNVNGGIYRHEFVASAVIDG MMNVQLSTGVPVLSAVLTPH RYRSSREHHEFFREHFMVKG VEAAAACITILAAREKIAA	I53-51B: 6, 7, 8, 9, 10, 11, 13, 18, 21, 27, 34, 38, 43, 48, 63, 67, 70, 85, 87, 101, 118, 125, 126, 129, 152, 153, 154
I52-03A SEQ ID NO: 11	MGHTKGPTPQQHDGSALRIG IVHARWNKTIIMPLLIGTIA KLECGVKASNIVVQSVPGS WELPIAVQRLYSASQLQTPS SGPSLSAGDLLGSSTDLTA LPTTTTASSTGPFDA LIAIGVLIKGETMHFEYIADSV SHGLMRVQLDTGVPVIFGVLT VLI DDQAKARAGVIEGSHNHG ED	I52-03A: 6, 9, 10, 11, 13, 15, 16, 26, 48, 69, 75, 76, 78, 79, 111, 125, 127, 142, 146, 159, 160, 161, 162, 171, 175, 193, 194, 196, 197, 199, 200

TABLE 2-continued		
Name	Amino Acid Sequence	Surface residues no tnearinterface
	WGLAAVEMGVRRRDWAAGKT E	
I52-03B SEQ ID NO: 12	MYEVDHADVYDLFYLG YAAEASDIADLVR SLLDVACGTGTHLEH GDTAGLELSE PDATLHQGDMRDFQ AVVSMFSSVGYLKT AVASFAEHLEPGGV WFPETFADGWVSAD RTVARVSHSVREGN HFTVADPGKGV TLFHQREYEAAFM YLEGGPSGRGLFV	I52-03B: 2, 3, 5, 6, 8, 15, 17, 20, 22, 23, 26, 27, 30, 33, 34, 35, 37, 38, 40, 54, 55, 57, 58, 59, 61, 62, 68, 70, 71, 74, 77, 78, 79, 81, 82, 84, 86, 87, 91, 96, 97, 98, 111, 127, 130, 131, 132, 141, 144, 145, 148, 150, 154, 157, 158, 159, 160, 161, 171, 172, 173, 174, 177, 187, 189, 192, 198, 199, 222, 223, 224, 236
I52-32A SEQ ID NO: 13	MGMKEKFVLIITHG LSGAEVIIGKQENV LGDNIEKVAKEVMR AEDKEIIIVVDLFG ALEMMKTFDVKVIT LVELLTSINVYDTT SKIGKDGIKVIEKSS	I52-32A: 3, 5, 15, 18, 30, 32, 35, 40, 41, 42, 44, 45, 65, 73, 79, 91, 103, 106, 109, 110, 111, 112, 114, 115, 118, 122, 123, 125, 126, 129, 131
I52-32B SEQ ID NO: 14	MKYDGSKLRI IIAALVAGAIKRLQ ENII IETVPGSFEL FVEKQKRLGKPLDA LIKGSTMHFEYICD MKLNFELGIPVIFG DEQAEARAGLIEGK DWGAAAVEMATKFN	I52-32B: 4, 6, 7, 9, 17, 32, 35, 42, 59, 63, 64, 66, 67, 68, 69, 70, 71, 73, 83, 85, 90, 106, 119, 120, 121, 122, 125, 131, 133, 134, 135, 136, 154
I52-33A SEQ ID NO: 15	MAVKGLGEVDQKY GILHARWNRKIIL LRLLFEGVKAENII SFELPYGSKLFVEK PLDAIIPIGVLIKGS YICDSTTHQLMKLN VIFGVLTCLTDEQA IEGKMHNHGEDWGA TKFN	I52-33A: 12, 14, 16, 17, 19, 26, 27, 46, 69, 73, 74, 76, 77, 78, 80, 81, 83, 93, 95, 100, 116, 129, 130, 131, 132, 145, 164
I52-33B SEQ ID NO: 16	MGANWYLDNESSR NADIAEVHRFLVLH GLAEVEVETESIST MLLRVLVFQVSKFP QLDMRPINNLAGAP PLTVSLRGKSHSYNA RLDERRFQVVTLEP DFDMVRAFNALRLV SLSVPGAVLIFTAR	I52-33B: 4, 6, 10, 20, 21, 23, 24, 31, 32, 34, 36, 39, 40, 42, 44, 46, 48, 56, 73, 77, 79, 81, 83, 85, 88, 89, 91, 92, 96, 97, 99, 101, 103, 109, 110, 111, 112, 114, 124, 125, 138, 140, 143, 158, 175
I32-06A SEQ ID NO: 17	MTDYIRDGSAIKALS AEADLRHIPQDLQRL HACGMVDVANDLAF AGRNALLAGAPILCD EGITRSRLPADNRV PSVPELAKKIGNTR LWLPHIEGSIVAIGN FRLFELLDAGAPKPA PVGFGAAESKDELA VPYVIVRGRRGGSAM NALASERE	I32-06A: 24, 26, 27, 41, 47, 50, 51, 56, 60, 63, 64, 67, 68, 77, 84, 85, 86, 91, 93, 98, 99, 100, 101, 102, 105, 108, 109, 114, 123, 124, 125, 127, 135, 142, 145, 148, 149, 152, 153, 169, 172, 173, 176, 177, 180, 187, 189
I32-06B SEQ ID NO: 18	MITVFGLKSKLAPR VIYSSLHLGLDIPK FLCLEKEDFYYPFD VIEINLMAGRSEET LLFIALERKLGIRAH IKEQPAHCWGFRGR DLDYDIYV	I32-06B: 8, 9, 10, 13, 14, 15, 16, 17, 20, 34, 36, 45, 46, 47, 50, 51, 53, 54, 57, 67, 70, 91, 93, 95, 105, 112

TABLE 2-continued

Name	Amino Acid Sequence	Surface residues no tnearinterface
I32-19A SEQ ID NO: 19	MGSDLQKLQRFSTCDISDGL LNVYNIPTGGYFPNLTAISP PQNSSIVGTAYTVLFAPIDD PRPAVNYIDSVPPNSILVLA LEPHLQSQFHPFIKITQAMY GGLMSTRAQYLKSNGTVVFG RIRDVDEHRTL NHPVFAYGV GSCAPKAVVKAVGTNVQLKI LTSDGVTQTICPGDYIAGDN NGIVRIPVQETDISKLVTYI EKSIEVDRLVSEAIKNGLPA KAAQTARRMVLKDYI	I32-19A: 3, 4, 6, 7, 9, 10, 25, 27, 36, 40, 42, 43, 44, 49, 58, 59, 61, 62, 63, 70, 72, 73, 74, 82, 84, 88, 89, 109, 110, 112, 126, 127, 129, 130, 132, 146, 155, 156, 157, 159, 166, 169, 172, 189, 190, 192, 194, 195, 198, 201, 204, 215, 232
I32-19B SEQ ID NO: 20	MSGMRVYLGADHAGYELKQA IIAFLKMTGHEPIDCGALRY DADDDYPAFCIAAATRTVAD PGSLGIVLGGSGNGEQIAAN KVPGARCALAWSVQTAALAR EHNNAQLIGIGGRMHTLEEA LRIVKAFVTPWSKAQRHQR RIDILAEYERTHEAPPVPGA PA	I32-19B: 4, 5, 31, 33, 38, 41, 42, 43, 55, 56, 59, 61, 62, 83, 93, 94, 101, 104, 113, 119, 129, 131, 134, 136, 137, 139, 140, 143, 144, 146, 147, 150, 152, 153, 156, 158, 159
I32-28A SEQ ID NO: 21	MGDDARIAAIGDVDELNSQI GVLLAEPLPDDVRAALSAIQ HDLFDLGELCIPGHAAITE DHLLRLALWL VHYNGQLPPL EEFILPGGARGAALAHVCRT VCRRRAERSIKALGASEPLNI APAAYVNLLSDLLFVLARVL NRAAGGADV LWDRTRAH	I32-28A: 4, 6, 7, 10, 14, 27, 30, 31, 33, 34, 41, 44, 45, 51, 52, 53, 54, 55, 56, 59, 76, 78, 79, 80, 81, 82, 83, 90, 103, 111, 115, 116, 131, 134, 142, 145, 147, 150
I32-28B SEQ ID NO: 22	MILSAEQSFTLRHPHGQAAA LAFVREPAAALAGVQRLRGL DSDGEQVWGELLVRVPLLGE VDLPFRSEIVRTPQGAELRP LTLTGERAWVAVSGQATAAE GGEMAFAFQFQAHLATPEAE GEGGAAFEVMVQAAAAGVTLL LVAMALPQGLAAGLPPA	I32-28B: 3, 4, 6, 8, 12, 15, 17, 18, 22, 26, 28, 32, 38, 39, 41, 43, 45, 46, 48, 50, 60, 66, 68, 71, 73, 74, 79, 81, 82, 83, 84, 86, 87, 95, 100, 103, 105, 109, 111, 113, 151, 152, 155, 156, 157
I53-40A.1 SEQ ID NO: 23	MTKKVGIVDTTFARVDMASA AILTLMESPNIKIIRKTVP GIKDLPVACKKLLEEEGCDI VMALGMPGKKEKDKVCAHEA SLGLMLAQLMTNKHIIIEVFV HEDEAKDDAELKILAARRAI EHALNVYYLLFKPEYLIRMA GKGLRQGFEDAGPARE	I53-40A: 3, 4, 31, 33, 35, 36, 37, 51, 54, 55, 56, 57, 59, 69, 70, 71, 74, 93, 103, 106, 118, 127, 128, 131, 132, 133, 134, 135, 138, 139, 142, 150, 153
I53-40B.1 SEQ ID NO: 24	MDDINNQLKRLKVIPVIAID NAEDIIPLGKVL AENGLPAA EITFRSSAAVKAIMLLRSAQ PEMLIGAGTILNGVQALAAK EAGADFVVS PGFNPNTVRAC QIIIGIDIVPGVNNPSTVEQA LEMGLTTLKFFPAEASGGIS MVKSLVGPYGD IRLMPTGGI TPDNIDNYLAIPQVLACGGT WMVDKKLVRNGEWDEIARLT REIVEQVNP	I53-40B: 2, 3, 7, 9, 10, 12, 20, 21, 23, 26, 27, 30, 34, 38, 45, 60, 62, 75, 85, 94, 95, 122, 124, 126, 134, 139, 143, 151, 153, 161, 163, 166, 167, 170, 172, 180, 184, 185, 186, 189, 190, 192, 193, 194, 195, 198, 201, 202, 205, 208, 209
I53-47A.1 SEQ ID NO: 25	MPIFTLNTNIKADDVPSDFL SLTSRLVGLILSKPGSYVAV HINTDQQLSFGGSTNPAAFG TLMSIGGIEPDKNRDHSAVL FDHLNAMLGIPKNRMYIHFV NLNGDDVGWNGTTF	I53-47A: 11, 13, 14, 17, 34, 36, 37, 45, 47, 54, 55, 56, 65, 69, 70, 71, 74, 91, 92, 93, 101, 103, 105, 109, 110, 112, 114
I53-47A.1NegT 2 SEQ ID NO: 26	MPIFTLNTNIKADDVPSDFL SLTSRLVGLILSEPGSYVAV HINIDQQLSFGGSTNPAAFG TLMSIGGIEPDKNEDHSAVL FDHLNAMLGIPKNRMYIHFV DLDGDDVGWNGTTF	I53-47A: 11, 13, 14, 17, 34, 36, 37, 45, 47, 54, 55, 56, 65, 69, 70, 71, 74, 91, 92, 93, 101, 103, 105, 109, 110, 112, 114

TABLE 2-continued

Name	Amino Acid Sequence	Surface residues no tnearinterface
I53-47B.1	MNQHSHKDHETVRIAVVRAR	I53-47B:
SEQ ID	WHADIVDACVEAFEIAMA	6, 7, 8, 9, 10, 11, 13, 18, 20, 21, 24,
NO: 27	GGDRFAVDVFDVPGAYE	43, 44, 51, 63, 67, 70, 85, 87, 101,
	HARTLAETGRYGAVLGTA	105, 122, 123, 124, 125, 126, 147, 152,
	VNGGIYRHEFVASAVIDG	153, 154
	NVQLDTGVPVLSAVLTPH	
	RDSDEHHRFFAAHFAVKG	
	AARACIEILNAREKIAA	
I53-47B.1NegT	MNQHSHKDHETVRIAVVRAR	I53-47B:
2	WHADIVDACVEAFEIAMA	6, 7, 8, 9, 10, 11, 13, 18, 20, 21, 24,
SEQ ID	GGDRFAVDVFDVPGAYE	43, 44, 51, 63, 67, 70, 85, 87, 101,
NO: 28	HARTLAETGRYGAVLGTA	105, 122, 123, 124, 125, 126, 147, 152,
	VDGGIYDHEFVASAVIDG	153, 154
	NVQLDTGVPVLSAVLTPH	
	EDSDEDHEFFAAHFAVKG	
	AARACIEILNAREKIAA	
I53-50A.1	MKMEELFKKHKIVAVLRANS	I53-50A:
SEQ ID	VEEAIEKAVAVFAGGVHL	4, 5, 6, 8, 9, 11, 17, 19, 23, 37, 46,
NO: 29	ITFTVPDADTVIKALSVL	47, 59, 74, 77, 78, 81, 94, 95, 98, 101,
	KGAIIGAGTVTSVEQCRKA	102, 103, 106, 119, 122, 126, 139, 142,
	ESGAEFIVSPHLDEEISQFC	145, 49, 150, 152, 160, 161, 162, 163,
	KEKGVFYMPGVMTPTELVKA	166, 169, 179, 183, 185, 188, 191, 192,
	MKLGHDIKLFPGEVVGPF	194, 198, 199
	VKAMKGPFPNVKFPVPTGG	
	LDNVCEWFKAGVLAVGVGD	
	LVKGDPEVREKAKKFVEKI	
	RGCTE	
I53-50A.1NegT	MKMEELFKKHKIVAVLRANS	I53-50A:
2	VEEAIEKAVAVFAGGVHL	4, 5, 6, 8, 9, 11, 17, 19, 23, 37, 46,
SEQ ID	ITFTVPDADTVIKALSVL	47, 59, 74, 77, 78, 81, 94, 95, 98, 101,
NO: 30	KGAIIGAGTVTSVEQCRKA	102, 103, 106, 119, 122, 126, 139, 142,
	ESGAEFIVSPHLDEEISQFC	145, 149, 150, 152, 160, 161, 162, 163,
	KEKGVFYMPGVMTPTELVKA	166, 169, 179, 183, 185, 188, 191, 192,
	MKLGHDIKLFPGEVVGPEF	194, 198, 199
	VEAMKGPFPNVKFPVPTGG	
	LDDVCEWFDAGVLAVGVGD	
	LVEGDPDEVREDAKEFVEEI	
	RGCTE	
I53-50A.1Post	MKMEELFKKHKIVAVLRANS	I53-50A:
1	VEEAIEKAVAVFAGGVHL	4, 5, 6, 8, 9, 11, 17, 19, 23, 37, 46,
SEQ ID	ITFTVPDADTVIKALSVL	47, 59, 74, 77, 78, 81, 94, 95, 98, 101,
NO: 31	KGAIIGAGTVTSVEQCRKA	102, 103, 106, 119, 122, 126, 139, 142,
	ESGAEFIVSPHLDEEISQFC	145, 419, 150, 152, 160, 161, 162, 163,
	KEKGVFYMPGVMTPTELVKA	166, 169, 179, 183, 185, 188, 191, 192,
	MKLGHDIKLFPGEVVGPF	194, 198, 199
	VKAMKGPFPNVKFPVPTGG	
	LDNVCKWFKAGVLAVGVGKA	
	LVKGGPDEVREKAKKFVKKI	
	RGCTE	
I53-50B.1	MNQHSHKDHETVRIAVVRAR	I53-50B:
SEQ ID	WHAIEIVDACVSAFEAAMR	6, 7, 8, 9, 10, 11, 13, 18, 20, 21, 34,
NO: 32	GGDRFAVDVFDVPGAYE	38, 93, 40, 43, 44, 48, 51, 63, 67, 70,
	HARTLAETGRYGAVLGTA	87, 101, 105, 118, 143, 147, 152, 153,
	VNGGIYRHEFVASAVIDG	154
	NVQLDTGVPVLSAVLTPH	
	RSDAHTLLFLALFAVKGME	
	AARACVEILAAREKIAA	
I53-50B.1NegT	MNQHSHKDHETVRIAVVRAR	I53-50B:
2	WHAIEIVDACVSAFEAAMR	6, 7, 8, 9, 10, 11, 13, 18, 20, 21, 34,
SEQ ID	GGDRFAVDVFDVPGAYE	38, 39, 40, 43, 44, 48, 51, 63, 67, 70,
NO: 33	HARTLAETGRYGAVLGTA	87, 101, 105, 118, 143, 147, 152, 153,
	VDGGIYDHEFVASAVIDG	154
	NVQLDTGVPVLSAVLTPH	
	EDSDADTLLFLALFAVKGME	
	AARACVEILAAREKIAA	
I53-50B.4Post	MNQHSHKDHETVRIAVVRAR	I53-50B:
1	WHAIEIVDACVSAFEAAMR	6, 7, 8, 9, 10, 11, 13, 18, 20, 21, 34,
	GGDRFAVDVFDVPGAYE	38, 39, 40, 43, 44, 48, 51, 63, 67, 70,

TABLE 2-continued

Name	Amino Acid Sequence	Surface residues no tnearinterface
SEQ ID NO: 34	HARTLAETGRYGAVLGTA FV VNGGIYRHEFVASAVINGMM NVQLNTGVPVLSAVLTPHNY DKSKAHTLLFLALFAVKGME AARACVEILAAREKIAA	87, 101, 105, 118, 143, 147, 152, 153, 154

[0032] In certain instances, the polypeptides of the present invention can also tolerate non-conservative substitutions. The isolated polypeptides may be produced recombinantly or synthetically, using standard techniques in the art. The isolated polypeptides of the invention can be modified in a number of ways, including but not limited to the ways described above, either before or after assembly of the nanostructures of the invention. As used throughout the present application, the term “polypeptide” is used in its broadest sense to refer to a sequence of subunit amino acids. The polypeptides of the invention may comprise L-amino acids, D-amino acids (which are resistant to L-amino acid-specific proteases in vivo), or a combination of D- and L-amino acids.

[0033] In another aspect, the invention provides nanostructures, comprising:

[0034] (a) a plurality of first assemblies, each first assembly comprising a plurality of identical first polypeptides, wherein the first polypeptides comprise the polypeptide of any embodiment or combination of embodiments of the first aspect of the invention; and

[0035] (b) a plurality of second assemblies, each second assembly comprising a plurality of identical second polypeptides, wherein the second polypeptides comprise the polypeptide of any embodiment or combination of embodiments of the first aspect of the invention, wherein the second polypeptide differs from the first polypeptide;

[0036] wherein the plurality of first assemblies non-covalently interact with the plurality of second assemblies to form a nanostructure.

[0037] As described in the examples that follow, a plurality (2, 3, 4, 5, 6, or more) of first polypeptides self-assemble to form a first assembly, and a plurality (2, 3, 4, 5, 6, or more) of second polypeptides self-assemble to form a second assembly. A plurality of these first and second assemblies then self-assemble non-covalently via the designed interfaces to produce the nanostructures of the invention. The designed interfaces on the polypeptides of the invention, resembling natural protein-protein interfaces with well-packed cores composed primarily of hydrophobic amino acid side chains surrounded by a periphery composed primarily of hydrophilic and charged side chains, rigidly orient the assemblies within the nanostructures formed by self-assembly. As will be understood by those of skill in the art, the interaction between the first assembly and the second assembly is a non-covalent protein-protein interaction. Any suitable non-covalent interaction(s) can drive self-interaction of the assemblies to form the nanostructure, including but not limited to one or more of electrostatic interactions, 7π -effects, van der Waals forces, hydrogen bonding, and hydrophobic effects. In various embodiments, pentamers, trimers, and dimers of the first or second assemblies assemble relative to each other such that their 5-fold, 3-fold,

and 2-fold symmetry axes are aligned along icosahedral 5-fold, 3-fold, and 2-fold symmetry axes, respectively.

[0038] In various other embodiments, the nanostructures are between about 20 nanometers (nm) to about 40 nm in diameter, with interior lumens between about 15 nm to about 32 nm across and pore sizes in the protein shells between about 1 nm to about 14 nm in their longest dimensions (FIG. 2). The nanostructures of the invention can be used for any suitable purpose, including but not limited to delivery vehicles, as the nanostructures can encapsulate molecules of interest and/or the first and/or second proteins can be modified to bind to molecules of interest (diagnostics, therapeutics, detectable molecules for imaging and other applications, etc.). The nanostructures of the invention are well suited for several applications, including vaccine design, targeted delivery of therapeutics, and bioenergy.

[0039] In various embodiments of the nanostructure of the invention, the first polypeptides and the second polypeptides comprise polypeptides with the amino acid sequence selected from the following pairs, or modified versions thereof (i.e.: permissible modifications as disclosed for the polypeptides of the invention: isolated polypeptides comprising an amino acid sequence that is at least 75% identical over its length, and identical at least at one identified interface position, to the amino acid sequence indicated by the SEQ ID NO.):

[0040] (i) SEQ ID NO:1 and SEQ ID NO:2 (I53-34A and I53-34B);

[0041] (ii) SEQ ID NO:3 and SEQ ID NO:4 (I53-40A and I53-40B);

[0042] (iii) SEQ ID NO:3 and SEQ ID NO:24 (I53-40A and I53-40B.1);

[0043] (iv) SEQ ID NO:23 and SEQ ID NO:4 (I53-40A.1 and I53-40B);

[0044] (v) SEQ ID NO:35 and SEQ ID NO:36 (I53-40A genus and I53-40B genus);

[0045] (vi) SEQ ID NO:5 and SEQ ID NO:6 (I53-47A and I53-B);

[0046] (vii) SEQ ID NO:5 and SEQ ID NO:27 (I53-47A and I53-47B.1);

[0047] (viii) SEQ ID NO:5 and SEQ ID NO:28 (I53-47A and I53-47B.1NegT2);

[0048] (ix) SEQ ID NO:25 and SEQ ID NO:6 (I53-47A.1 and I53-47B);

[0049] (x) SEQ ID NO:25 and SEQ ID NO:27 (I53-47A.1 and I53-47B.1);

[0050] (xi) SEQ ID NO:25 and SEQ ID NO:28 (I53-47A.1 and I53-47B.1NegT2);

[0051] (xii) SEQ ID NO:26 and SEQ ID NO:6 (I53-47A.1NegT2 and I53-47B);

[0052] (xiii) SEQ ID NO:26 and SEQ ID NO:27 (I53-47A.1NegT2 and I53-47B.1);

[0053] (xiv) SEQ ID NO:26 and SEQ ID NO:28 (I53-47A.1NegT2 and I53-47B.1NegT2);

[0054] (xv) SEQ ID NO:37 and SEQ ID NO:38 (I53-47A genus and I53-47B genus);

[0055] (xvi) SEQ ID NO:7 and SEQ ID NO:8 (I53-50A and I53-50B);

[0056] (xvii) SEQ ID NO:7 and SEQ ID NO:32 (I53-50A and I53-50B.1);

[0057] (xix) SEQ ID NO:7 and SEQ ID NO:33 (I53-50A and I53-50B.1NegT2);

[0058] (xx) SEQ ID NO:7 and SEQ ID NO:34 (I53-50A and I53-50B.4PosT1);

[0059] (xxi) SEQ ID NO:29 and SEQ ID NO:8 (I53-50A.1 and I53-50B);

[0060] (xxii) SEQ ID NO:29 and SEQ ID NO:32 (I53-50A.1 and I53-50B.1);

[0061] (xxiii) SEQ ID NO:29 and SEQ ID NO:33 (I53-50A.1 and I53-50B.1NegT2);

[0062] (xxiv) SEQ ID NO:29 and SEQ ID NO:34 (I53-50A.1 and I53-50B.4PosT1);

[0063] (xxv) SEQ ID NO:30 and SEQ ID NO:8 (I53-50A.1NegT2 and I53-50B);

[0064] (xxvi) SEQ ID NO:30 and SEQ ID NO:32 (I53-50A.1NegT2 and I53-50B.1);

[0065] (xxvii) SEQ ID NO:30 and SEQ ID NO:33 (I53-50A.1NegT2 and I53-50B.1NegT2);

[0066] (xxviii) SEQ ID NO:30 and SEQ ID NO:34 (I53-50A.1NegT2 and I53-50B.4PosT1);

[0067] (xxix) SEQ ID NO:31 and SEQ ID NO:8 (I53-50A.1PosT1 and I53-50B);

[0068] (xxx) SEQ ID NO:31 and SEQ ID NO:32 (I53-50A.1PosT1 and I53-50B.1);

[0069] (xxxi) SEQ ID NO:31 and SEQ ID NO:33 (I53-50A.1PosT1 and I53-50B.1NegT2);

[0070] (xxxii) SEQ ID NO:31 and SEQ ID NO:34 (I53-50A.1PosT1 and I53-50B.4PosT1);

[0071] (xxxiii) SEQ ID NO:39 and SEQ ID NO:40 (I53-50A genus and I53-50B genus);

[0072] (xxxiv) SEQ ID NO:9 and SEQ ID NO:10 (I53-51A and I53-51B);

[0073] (xxxv) SEQ ID NO:11 and SEQ ID NO:12 (I52-03A and I52-03B);

[0074] (xxxvi) SEQ ID NO:13 and SEQ ID NO:14 (I52-32A and I52-32B);

[0075] (xxxv) SEQ ID NO:15 and SEQ ID NO:16 (I52-33A and I52-33B);

[0076] (xxxvi) SEQ ID NO:17 and SEQ ID NO:18 (I32-06A and I32-06B);

[0077] (xxxvii) SEQ ID NO:19 and SEQ ID NO:20 (I32-19A and I32-19B);

[0078] (xxxviii) SEQ ID NO:21 and SEQ ID NO:22 (I32-28A and I32-28B); and

[0079] (xxxix) SEQ ID NO:23 and SEQ ID NO:24 (I53-40A.1 and I53-40B.1).

[0080] In one embodiment, the nanostructure has icosahedral symmetry. In this embodiment, the nanostructure may comprise 60 copies of the first polypeptide and 60 copies of the second polypeptide. In one such embodiment, the number of identical first polypeptides in each first assembly is different than the number of identical second polypeptides in each second assembly. For example, in one embodiment, the nanostructure comprises twelve first assemblies and twenty second assemblies; in this embodiment, each first assembly may, for example, comprise five copies of the identical first polypeptide, and each second assembly may, for example, comprise three copies of the identical second polypeptide. In

another embodiment, the nanostructure comprises twelve first assemblies and thirty second assemblies; in this embodiment, each first assembly may, for example, comprise five copies of the identical first polypeptide, and each second assembly may, for example, comprise two copies of the identical second polypeptide. In a further embodiment, the nanostructure comprises twenty first assemblies and thirty second assemblies; in this embodiment, each first assembly may, for example, comprise three copies of the identical first polypeptide, and each second assembly may, for example, comprise two copies of the identical second polypeptide. All of these embodiments are capable of forming synthetic nanomaterials with regular icosahedral symmetry. In various further embodiments, oligomeric states of the first and second polypeptides are as follows:

[0081] I53-34A: trimer+I53-34B: pentamer;

[0082] I53-40A: pentamer+I53-40B: trimer;

[0083] I53-47A: trimer+I53-47B: pentamer;

[0084] I53-50A: trimer+I53-50B: pentamer;

[0085] I53-51A: trimer+I53-51B: pentamer;

[0086] I32-06A: dimer+I32-06B: trimer;

[0087] I32-19A: trimer+I32-19B: dimer;

[0088] I32-28A: trimer+I32-28B: dimer;

[0089] I52-03A: pentamer+I52-03B: dimer;

[0090] I52-32A: dimer+I52-32B: pentamer; and

[0091] I52-33A: pentamer+I52-33B: dimer.

[0092] As disclosed in the examples that follow, the nanostructures form spontaneously when appropriate polypeptide pairs are co-expressed in *E. coli* cells, yielding milligram quantities of purified material per liter of cell culture using standard methods of immobilized metal-affinity chromatography and gel filtration. When a poly-histidine purification tag is appended to just one of the two distinct polypeptide subunits (i.e.: the first and second polypeptides) comprising each nanostructure, the other subunit is found to co-purify with the tagged subunit.

[0093] In one embodiment, the nanostructure further comprises a cargo within the nanostructure. As used herein, a “cargo” is any compound or material that can be incorporated on and/or within the nanostructure. For example, polypeptide pairs suitable for nanostructure self-assembly can be expressed/purified independently; they can then be mixed in vitro in the presence of a cargo of interest to produce the nanostructure comprising a cargo. This feature, combined with the protein nanostructures’ large lumens and relatively small pore sizes, makes them well suited for the encapsulation of a broad range of cargo including, but not limited to, small molecules, nucleic acids, polymers, and other proteins. In turn, the protein nanostructures of the present invention could be used for many applications in medicine and biotechnology, including targeted drug delivery and vaccine design. For targeted drug delivery, targeting moieties could be fused or conjugated to the protein nanostructure exterior to mediate binding and entry into specific cell populations and drug molecules could be encapsulated in the cage interior for release upon entry to the target cell or sub-cellular compartment. For vaccine design, antigenic epitopes from pathogens could be fused or conjugated to the cage exterior to stimulate development of adaptive immune responses to the displayed epitopes, with adjuvants and other immunomodulatory compounds attached to the exterior and/or encapsulated in the cage interior to help tailor the type of immune response generated for each pathogen. The polypeptide components may be modified as noted above. In

one non-limiting example, the polypeptides can be modified, such as by introduction of various cysteine residues at defined positions to facilitate linkage to one or more antigens of interest as cargo, and the nanostructure could act as a scaffold to provide a large number of antigens for delivery as a vaccine to generate an improved immune response. Other modifications of the polypeptides as discussed above may also be useful for incorporating cargo into the nanostructure.

[0094] In certain embodiments, the nanostructures may comprise one or more peptides configured to bind or fuse with desired immunogens. In certain further embodiments, the nanostructure comprises one or more copies of variants designed to form a nanostructure of the trimeric proteins 1WOZ or 1WA3 (PDB ID codes), which have been demonstrated to be suitable for fusion with the trimeric HIV immunogen, BG505 SOSIP (4-6). Such nanostructures could be used as scaffolds for the design of an HIV vaccine capable of inducing protective immune responses against the virus. In another embodiment, the nanostructures of the present invention could be useful as scaffolds for the attachment of enzymes on the interior and/or exterior of the cages. Such enzymes confer on the nanostructure the ability to catalyze biochemical pathways or other reactions. Such patterning has been shown to be important in natural systems in order to increase local substrate concentrations, sequester toxic intermediates, and/or reduce the rates of undesirable side reactions (7, 8). In another embodiment, the cargo may comprise a detectable cargo. For example, the nanostructures of the present invention could also be useful as single-cell or single-molecule imaging agents. The materials are large enough to be identified in cells by electron microscopy, and when tagged with fluorophores they are readily detectable by light microscopy. This feature makes them well-suited to the task of correlating images of the same cells taken by light microscopy and electron microscopy.

[0095] In another aspect, the present invention provides isolated nucleic acids encoding a protein of the present invention. The isolated nucleic acid sequence may comprise RNA or DNA. As used herein, “isolated nucleic acids” are those that have been removed from their normal surrounding nucleic acid sequences in the genome or in cDNA sequences. Such isolated nucleic acid sequences may comprise additional sequences useful for promoting expression and/or purification of the encoded protein, including but not limited to polyA sequences, modified Kozak sequences, and sequences encoding epitope tags, export signals, and secretory signals, nuclear localization signals, and plasma membrane localization signals. It will be apparent to those of skill in the art, based on the teachings herein, what nucleic acid sequences will encode the proteins of the invention.

[0096] In a further aspect, the present invention provides recombinant expression vectors comprising the isolated nucleic acid of any embodiment or combination of embodiments of the invention operatively linked to a suitable control sequence. “Recombinant expression vector” includes vectors that operatively link a nucleic acid coding region or gene to any control sequences capable of effecting expression of the gene product. “Control sequences” operatively linked to the nucleic acid sequences of the invention are nucleic acid sequences capable of effecting the expression of the nucleic acid molecules. The control sequences need not be contiguous with the nucleic acid sequences, so long as

they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the nucleic acid sequences and the promoter sequence can still be considered “operably linked” to the coding sequence. Other such control sequences include, but are not limited to, polyadenylation signals, termination signals, and ribosome binding sites. Such expression vectors can be of any type known in the art, including but not limited to plasmid and viral-based expression vectors. The control sequence used to drive expression of the disclosed nucleic acid sequences in a mammalian system may be constitutive (driven by any of a variety of promoters, including but not limited to, CMV, SV40, RSV, actin, EF) or inducible (driven by any of a number of inducible promoters including, but not limited to, tetracycline, ecdysone, steroid-responsive). The construction of expression vectors for use in transfecting prokaryotic cells is also well known in the art, and thus can be accomplished via standard techniques. (See, for example, Sambrook, Fritsch, and Maniatis, in: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989; *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E. J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, Tex.). The expression vector must be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. In a preferred embodiment, the expression vector comprises a plasmid. However, the invention is intended to include other expression vectors that serve equivalent functions, such as viral vectors.

[0097] In another aspect, the present invention provides host cells that have been transfected with the recombinant expression vectors disclosed herein, wherein the host cells can be either prokaryotic or eukaryotic. The cells can be transiently or stably transfected. Such transfection of expression vectors into prokaryotic and eukaryotic cells can be accomplished via any technique known in the art, including but not limited to standard bacterial transformations, calcium phosphate co-precipitation, electroporation, or liposome mediated-, DEAE dextran mediated-, polycationic mediated-, or viral mediated transfection. (See, for example, *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press; *Culture of Animal Cells: A Manual of Basic Technique*, 2nd Ed. (R. I. Freshney, 1987, Liss, Inc. New York, N.Y.). A method of producing a polypeptide according to the invention is an additional part of the invention. The method comprises the steps of (a) culturing a host according to this aspect of the invention under conditions conducive to the expression of the polypeptide, and (b) optionally, recovering the expressed polypeptide.

[0098] In a further aspect, the present invention provides kits comprising:

[0099] (a) one or more of the isolated polypeptides, polypeptide assemblies, or nanostructures of the invention;

[0100] (b) one or more recombinant nucleic acids of the invention;

[0101] (c) one or more recombinant expression vectors comprising recombinant nucleic acids of the invention; and/or

[0102] (d) one or more recombinant host cell, comprising recombinant expression vectors of the invention.

[0103] In yet a further aspect, the present invention provides methods of using the nanostructures of the present

invention. In cases where both polypeptides comprising an assembly are capable of independent expression and purification, this enables control over assembly through mixing of purified components in vitro. This feature, combined with the nanostructures' large lumens and relatively small pore sizes, makes them well suited for the encapsulation of a broad range of other materials including small molecules, nucleic acids, polymers, and other proteins, as discussed above. In turn, the nanostructures of the present invention could be used for many applications in medicine and biotechnology, including targeted drug delivery and vaccine design. For targeted drug delivery, targeting moieties could be fused or conjugated to the nanostructure exterior to mediate binding and entry into specific cell populations and drug molecules could be encapsulated in the cage interior for release upon entry to the target cell or sub-cellular compartment. For vaccine design, antigenic epitopes from pathogens could be fused or conjugated to the nanostructure exterior to stimulate development of adaptive immune responses to the displayed epitopes, with adjuvants and other immunomodulatory compounds attached to the exterior and/or encapsulated in the cage interior to help tailor the type of immune response generated for each pathogen. Other uses will be clear to those of skill in the art based on the disclosure relating to polypeptide modifications, nanostructure design, and cargo incorporation.

Examples

[0104] Methods of production: The icosahedral materials disclosed herein (amino acid sequences provided in Table 1), which comprise possible embodiments of the present invention, were produced as follows. The initial sequences and structures for the design process were derived from pentameric, trimeric, and dimeric crystal structures from the Protein Data Bank (PDB), along with a small number of crystal structures of de novo designed proteins not yet deposited in the PDB.

[0105] The PDB Accession numbers for the wild type scaffold proteins related to the exemplary polypeptides of the invention are as follows:

- [0106]** SEQ ID NO:1 (I53-34A): 2yw3;
- [0107]** SEQ ID NO:2 (I53-34B): 2b98;
- [0108]** SEQ ID NO:3 (I53-40A): 2b98;
- [0109]** SEQ ID NO:4 (I53-40B): 4e38;
- [0110]** SEQ ID NO:5 (I53-47A): 1hfo;
- [0111]** SEQ ID NO:6 (I53-47B): 2obx;
- [0112]** SEQ ID NO:7 (I53-50A): 1wa3;
- [0113]** SEQ ID NO:8 (I53-50B): 2obx;
- [0114]** SEQ ID NO:9 (I53-51A): 1woz;
- [0115]** SEQ ID NO:10 (I53-51B): 2obx;
- [0116]** SEQ ID NO:11 (I52-03A): 1c41;
- [0117]** SEQ ID NO:12 (I52-03B): 3bxo;
- [0118]** SEQ ID NO:13 (I52-32A): 31fh;
- [0119]** SEQ ID NO:14 (I52-32B): 2jfb;
- [0120]** SEQ ID NO:15 (I52-33A): 2jfb;
- [0121]** SEQ ID NO:16 (I52-33B): 3q34;
- [0122]** SEQ ID NO:17 (I32-06A): 3e7d;
- [0123]** SEQ ID NO:18 (I32-06B): 1mww;
- [0124]** SEQ ID NO:19 (I32-19A): 2c5q;
- [0125]** SEQ ID NO:20 (I32-19B): 2vvp;
- [0126]** SEQ ID NO:21 (I32-28A): 2zhz; and
- [0127]** SEQ ID NO:22 (I32-28B): 3nqn.
- [0128]** 15,552 pairs of pentamers and trimers, 50,400 pairs of pentamers and dimers, and 344,825 pairs of trimers and

dimers were arranged in icosahedral symmetry with the 5-fold symmetry axes of the pentamers, 3-fold symmetry axes of the trimers, and 2-fold symmetry axes of the dimers aligned along the 5-fold, 3-fold, and 2-fold icosahedral symmetry axes, respectively. While maintaining perfect icosahedral symmetry, rotations and translations along these axes were sampled to identify configurations predicted to be suitable for protein-protein interface design. In total, 68,983 153, 35,468 152, and 177,252 132 configurations were designed, yielding 71 pairs of 153 protein sequences, 44 pairs of I52 protein sequences, and 68 pairs of I32 protein sequences predicted to fold and assemble into the modeled icosahedral complexes.

[0129] Genes encoding the 71 pairs of I53 sequences were synthesized and cloned into a variant of the pET29b expression vector (Novagen, Inc.) between the NdeI and XhoI endonuclease restriction sites. Genes encoding the 44 pairs of I52 sequences and 68 pairs of I32 sequences were synthesized and cloned into a variant of the pET28b expression vector (Novagen, Inc.) between the NcoI and XhoI endonuclease restriction sites.

[0130] The two protein coding regions in each DNA construct are connected by an intergenic region. The intergenic region in the I53 designs was derived from the pETDuet-1 vector (Novagen, Inc.) and includes a stop codon, T7 promoter/lac operator, and ribosome binding site.

[0131] The intergenic region in the I52 and I32 designs only includes a stop codon and ribosome binding site. The sequences of the 153, 152 and I32 intergenic regions are as follows:

[0132] I53 intergenic region DNA sequence:

(SEQ ID NO: 41)

```

5' -
TAATGCTTAAGTCGAACAGAAAGTAATC
GTATTGTACACGGCCGCATAATCGAAAT
TAATACGACTCACTATAGGGGAATTGTG
AGCGGATAACAATTCCCCATCTTAGTAT
ATTAGTTAAGTATAAGAAGGAGATATAC
TT-3'

```

[0133] I52 intergenic region DNA sequence:

(SEQ ID NO: 42)

```

5' - TAAAGAAGGAGATATCAT-3'

```

[0134] I32 intergenic region DNA sequence:

(SEQ ID NO: 43)

```

5' - TGAGAAGGAGATATCAT-3'

```

[0135] The constructs for the I53 protein pairs thus possess the following set of elements from 5' to 3': NdeI restriction site, upstream gene, intergenic region, downstream gene, XhoI restriction site. The constructs for the I52 and I32 protein pairs possess the following set of elements from 5' to 3': NcoI restriction site, upstream gene, intergenic region, downstream gene, XhoI restriction site. In each case, the upstream genes encode components denoted with the suffix "A"; the downstream genes encode the "B" compo-

nents (Table 1). This allows for co-expression of the designed protein pairs in which both the upstream and downstream genes have their own ribosome binding site, and in the case of the I53 designs, both genes also have their own T7 promoter/lac operator.

[0136] For purification purposes, each co-expression construct includes a 6x-histidine tag (HHHHHH) appended to the N- or C-terminus of one of the two protein coding regions.

[0137] Expression plasmids were transformed into BL21 (DE3) *E. coli* cells. Cells were grown in LB medium supplemented with 50 mg L⁻¹ of kanamycin (Sigma) at 37° C. until an OD600 of 0.8 was reached. Protein expression was induced by addition of 0.5 mM isopropyl-thio- β -D-galactopyranoside (Sigma) and allowed to proceed for either 5 h at 22° C. or 3 h at 37° C. before cells were harvested by centrifugation.

[0138] The designed proteins were first screened for soluble expression and co-purification at small scale from 2 to 4 mL cultures by nickel affinity chromatography using His MultiTrap® FF nickel-coated filter plates (GE Healthcare). Purification products were analyzed by SDS-PAGE to identify those containing species near the expected molecular weight of both protein subunits (indicating co-purification). Those found to contain both subunits were subsequently subjected to native (non-denaturing) PAGE to identify slow migrating species further indicating assembly to higher order materials. Those designs appearing to co-purify and yielding slowly migrating species by native PAGE were subsequently expressed at larger scale (1 to 12 liters of culture) and purified by nickel affinity chromatography via gravity columns with nickel-NTA resin (Qiagen) or His-Trap® HP columns (GE Healthcare). Fractions containing the designed proteins were pooled, concentrated using centrifugal filter devices (Sartorius Stedim Biotech), and further purified on a Superose® 6 10/300 gel filtration column (GE Healthcare).

[0139] The purified proteins were analyzed by size exclusion chromatography using a Superose® 6 10/300 column to assess their assembly states. For each of the exemplary proteins described here, major peaks were observed in the chromatograms near elution volumes of 8.5 to 12 mL, which correspond well with the expected elution volumes for the designed 120-subunit icosahedral nanostructures. Within this set of exemplary proteins, the relative elution volumes correspond with the physical dimensions of the computational design models of the nanostructures, that is, proteins designed to assemble into relatively larger nanostructures yielded peaks at earlier elution volumes while those designed to assemble into relatively smaller nanostructures yielded peaks at later elution volumes. In some cases, smaller secondary peaks were observed at slightly earlier elution volumes than the predominant peak, suggesting transient or low-affinity dimerization of the nanostructures.

[0140] Gel filtration fractions containing pure protein in the desired assembly state were analyzed by negative stain electron microscopy as described previously (2). Electron micrographs showing fields of particles of the expected size and shape have been obtained for 10 of the nanostructures. In one case (I32-19), the nanostructure appears to be unstable in the conditions encountered during grid preparation, precluding visualization by electron microscopy.

[0141] To further validate the structures of our materials, small angle X-ray scattering (SAXS) data was obtained for

several of the designed nanostructures. Scattering measurements were performed at the SIBYLS® 12.3.1 beamline at the Advanced Light Source, LBNL, on 20 microliter samples loaded into a helium-purged sample chamber (10). Data were collected on gel filtration fractions and samples concentrated ~2x-10x from individual fractions, with the gel filtration buffer and concentrator eluates used for buffer subtraction. Sequential exposures ranging from 0.5 to 5 seconds were taken at 12 keV to maximize signal to noise, with visual checks for radiation-induced damage to the protein. The FOXS® algorithm (11, 12) was then used to calculate scattering profiles from our design models and fit them to the experimental data. The major features of the I53-34, I53-40, I53-47, I53-50, I52-03, I52-32, I52-33, I32-06, I32-19, and I32-28 design models were all found to match well with the experimental data, supporting the conclusion that the nanostructures assemble to the intended assembly state and three-dimensional configuration in solution. Graphs of the log of the scattering intensity, $I(q)$, as a function of scattering angle, q , show multiple large dips in the scattering intensity in the low q region between 0.015 Å⁻¹ and 0.15 Å⁻¹, each of which is closely recapitulated in the theoretical profiles calculated from the design models. Although the I53-51 design model was not found to match well with the SAXS data, this appears likely to be due to low stability of the designed material, which caused it to be primarily unassembled at the concentrations used for the SAXS measurements; this result is consistent with our findings from gel filtration of I53-51, in which significant peaks were observed corresponding to the unassembled pentamers and trimers in addition to the presumed 120-subunit assembly peak.

[0142] Using the Rosetta macromolecular modeling suite, the computational models of designed I53 materials were redesigned by allowing optimization of the identities of relatively exposed residues (defined as having a solvent accessible surface area of greater than 20 square Ångstroms), excepting polar residues (Aspartate, Glutamate, Histidine, Lysine, Asparagine, Glutamine, and Arginine) and residues near the designed protein-protein interfaces between the pentameric and trimeric components. Mutations that resulted in losses of significant atomic packing interactions or side chain-backbone hydrogen bonds were discarded. A position-specific scoring matrix (PSSM) based on homologous protein sequences was used to augment the Rosetta scorefunction to favor residues that appear frequently at a given position in homologous proteins, a design approach referred to as consensus protein design (9). Multiple design trajectories were performed with varying weights on the contribution of the PSSM, and mutations to polar residues that appeared favorable across all design trajectories were selected for inclusion in the variant protein. These variants were designated by the addition of “0.1” to the end of their names (e.g., I53-50A. 1).

[0143] The Rosetta macromolecular modeling suite was used to mutate manually selected amino acid positions to charged amino acids in order to generate variant nanoparticles featuring highly positively or negatively charged interior surfaces. To generate negatively charged nanoparticles (denoted by the letters “Neg” in their names), mutations were limited to either Aspartate or Glutamate. To generate positively charged nanoparticles (denoted by the letters “Pos” in their names), mutations were limited to either Arginine or Lysine. Relevant score metrics for each muta-

tion were independently assessed, and favorable mutations were sorted into two tiers based on their scores. Two new nanoparticle variants sequences were then designed for each individual protein for each type of charge, one including only the Tier 1 mutations (named “T1”) and the other including both the Tier 1 and Tier 2 mutations (named “T2”). In most cases, the charged mutations were incorporated into the consensus redesign variants described above.

[0144] Genes encoding the I53 “0.1” and charged variant proteins were synthesized and cloned into the pET29b expression vector (Novagen, Inc.) between the NdeI and XhoI endonuclease restriction sites. Constructs were produced in two formats. In the first, the two proteins were encoded in a bicistronic arrangement on a single expression plasmid as described above for co-expression in *E. coli*. In the second, each protein component (i.e., the pentameric component and the trimeric component) were cloned individually into pET29b for expression in the absence of the other component.

[0145] For purification purposes, each co-expression construct included a 6x-histidine tag (HHHHHH) appended to the N- or C-terminus of one of the two protein coding regions. Similarly, each individual expression construct included a 6x-histidine tag appended to the N- or C-terminus of the protein coding region.

[0146] The “0.1” and charged variant proteins were expressed and purified as described above with two differences. First, expression at 18° C. was evaluated in addition to expression at 37° C. at small scale for all variants, and, in some cases, expression at 18° C. was used to produce the proteins at multi-liter scale. Second, for some variants, the detergent 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) was included in all purification buffers at a concentration of 0.75% weight/volume to prevent protein aggregation.

[0147] After purification of individually expressed protein components, pairs of components designed to co-assemble into a nanoparticle (e.g., I53-40.1A and I53-40.1B) were mixed in equimolar amounts in buffer and allowed to incubate at room temperature for 1-24 hours, a procedure we refer to as “in vitro assembly.” For assemblies including charged components, the buffer included 500 mM NaCl; in all other cases the buffer included 150 mM NaCl. The mixtures were fractionated and analyzed on a Superose® 6 10/300 gel filtration column (GE Healthcare), and fractions were analyzed by SDS-PAGE to determine the protein contents of each elution peak.

[0148] In one exemplary embodiment, the I53-40.1A and I53-40.1B protein variants, based off of I53-40A and I53-40B, respectively, were constructed by consensus protein design, in which multiple sequence alignments from protein families related to each protein subunit were used to guide the selection of amino acid residues at surface-exposed positions. The variant proteins were found to be more stable and soluble when purified independently than the original proteins, a property that enabled the formation of the designed nanostructure by simply mixing solutions containing the purified components in physiological buffers in a 1:1 molar ratio. The addition of 0.75% CHAPS, a zwitterionic detergent, to the buffer was found to further increase the stability and solubility of I53-40.1A and was therefore included during the purification of the protein prior to in vitro assembly. Size exclusion chromatograms from a run analyzing the mixed solution containing both components

on a Superose 6 column revealed a single major peak at the elution volume expected for the 120-subunit designed icosahedral nanostructure. Analysis of the peak fractions by SDS-PAGE revealed bands at the expected molecular weight for the first and second polypeptides of the nanostructure in an apparent 1:1 stoichiometric ratio. The data demonstrate that when mixed, the two components co-assemble to the 120-subunit designed icosahedral nanostructure.

[0149] In another exemplary embodiment, the I53-47A.1, I53-47B.1, I53-50A.1, and I53-50B.1 protein variants, based off of I53-47A, I53-47B, I53-50A, and I53-50B, respectively, were constructed by consensus protein design, in which multiple sequence alignments from protein families related to each protein subunit were used to guide the selection of amino acid residues at surface-exposed positions. The variant proteins were found to be more stable and soluble when purified independently than the original proteins, a property that enabled the formation of the designed nanostructure by simply mixing solutions containing the purified components in physiological buffers in a 1:1 molar ratio, a process referred to as in vitro assembly. The addition of 0.75% CHAPS, a zwitterionic detergent, to the buffer was found to further increase the stability and solubility of I53-47B.1 and I53-50B.1 and was therefore included during the purification of the proteins prior to in vitro assembly. Size exclusion chromatograms from a run analyzing the mixed solution containing both I53-47A.1 and I53-47B.1 on a Superose 6 column revealed a major peak at the elution volume expected for the 120-subunit designed icosahedral nanostructure as well as a smaller secondary peak at a later elution volume. Analysis of the peak fractions corresponding to the 120-subunit nanostructure by SDS-PAGE revealed bands at the expected molecular weight for the first and second polypeptides of the nanostructure in an apparent 1:1 stoichiometric ratio. Analysis of the secondary peak at the later elution volume revealed that this peak comprises only the trimeric subunit, suggesting that the in vitro assembly mixture actually contained an excess of this polypeptide. Similarly, size exclusion chromatograms from a run analyzing the mixed solution containing both I53-50A.1 and I53-50B.1 on a Superose 6 column revealed a peak at the elution volume expected for the 120-subunit designed icosahedral nanostructure as well as two secondary peaks at later elution volumes. Analysis of the peak fractions corresponding to the 120-subunit nanostructure by SDS-PAGE revealed bands at the expected molecular weight for the first and second polypeptides of the nanostructure in an apparent 1:1 stoichiometric ratio. Analysis of the secondary peaks at the later elution volumes revealed that the first of the two comprises only the pentameric subunit, while the second of the two comprises only the trimeric subunit, suggesting that for this pair of proteins, in vitro assembly is somewhat inefficient. Together, the data demonstrate that when mixed, the two components of each nanostructure (i.e., I53-47A.1 and I53-47B.1 or I53-50A.1 and I53-50B.1) co-assemble to the 120-subunit designed icosahedral nanostructures.

[0150] In another exemplary embodiment, the protein variants I53-47A.1NegT2, I53-47B.1NegT2, I53-50A.1NegT2, and I53-50B.1NegT2, based off of I53-47A.1, I53-47B.1, I53-50A.1, and I53-50B.1, respectively, bear mutations that introduce additional negatively charged amino acid residues (i.e., Aspartate and Glutamate) on their surfaces such that the nanostructures formed through the

assembly of these proteins have highly charged interior surfaces. After the two independently purified proteins I53-47A.1NegT2 and I53-47B.1NegT2 were mixed together in an in vitro assembly reaction in a buffer with a concentration of 150 mM NaCl, no assembly was observed when the mixture was analyzed on a Superose 6 size exclusion chromatography column; only unassembled I53-47A.1NegT2 and I53-47B.1NegT2 proteins eluted from the column. In contrast, if the in vitro assembly reaction was performed in the presence of 0.5 M NaCl, robust assembly to the designed nanostructure was observed, with some remaining unassembled proteins eluting later as smaller secondary elution peaks. Similarly, after the two independently purified proteins I53-50A.1NegT2 and I53-50B.1NegT2 were mixed together in an in vitro assembly reaction in a buffer with a concentration of 150 mM NaCl, no assembly was observed when the mixture was analyzed on a Superose® 6 size exclusion chromatography column; only unassembled I53-50A.1NegT2 and I53-50B.1NegT2 proteins eluted from the column. In contrast, if the in vitro assembly reaction was performed in the presence of 0.5 M NaCl, assembly to the designed nanostructure was observed, with some remaining unassembled proteins eluting later. Together, the data demonstrate that when mixed, the two components of each highly charged 120-subunit designed icosahedral nanostructure assemble to the target structure only in the presence of high ionic strength.

[0151] In order to package nucleic acids, pairs of individually purified protein components designed to co-assemble into a nanoparticle were combined with single-stranded DNA (ssDNA) in buffer and allowed to incubate overnight. ssDNA was present at a final concentration of 26 ng/μL (200 pM) for 400 nucleotide (nt) strands, and 35.2 ng/μL (66.7 pM) for 1600 nt strands. Individual protein components were added at final equimolar concentrations ranging from 2-12 and the final NaCl concentration was 150 mM. After overnight incubation, samples were either analyzed by electrophoresis on a 1% agarose gel or DNase I was added to a final concentration of 25 pg/mL and incubated for one hour at room temperature before electrophoresis. Gels were stained with SybrGold® (ThermoFisher Scientific) and imaged to visualize nucleic acid, and were subsequently stained with GelCode® Blue (Pierce) and imaged again to visualize protein.

[0152] The above definitions and explanations are meant and intended to be controlling in any future construction unless clearly and unambiguously modified in the following examples or when application of the meaning renders any construction meaningless or essentially meaningless. In cases where the construction of the term would render it meaningless or essentially meaningless, the definition should be taken from Webster's Dictionary, 3rd Edition or a dictionary known to those of skill in the art, such as the Oxford Dictionary of Biochemistry and Molecular Biology (Ed. Anthony Smith, Oxford University Press, Oxford, 2004).

[0153] The above description provides specific details for a thorough understanding of, and enabling description for, embodiments of the disclosure. However, one skilled in the art will understand that the disclosure may be practiced without these details. In other instances, well-known structures and functions have not been shown or described in detail to avoid unnecessarily obscuring the description of the embodiments of the disclosure. The description of embodi-

ments of the disclosure is not intended to be exhaustive or to limit the disclosure to the precise form disclosed. While specific embodiments of, and examples for, the disclosure are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the disclosure, as those skilled in the relevant art will recognize.

[0154] Aspects of the disclosure can be modified, if necessary, to employ the systems, functions and concepts of the above references and application to provide yet further embodiments of the disclosure. These and other changes can be made to the disclosure in light of the detailed description.

[0155] Specific elements of any of the foregoing embodiments can be combined or substituted for elements in other embodiments. Furthermore, while advantages associated with certain embodiments of the disclosure have been described in the context of these embodiments, other embodiments may also exhibit such advantages, and not all embodiments need necessarily exhibit such advantages to fall within the scope of the disclosure.

[0156] The above detailed description describes various features and functions of the disclosed systems, devices, and methods with reference to the accompanying figures. In the figures, similar symbols typically identify similar components, unless context dictates otherwise. The illustrative embodiments described in the detailed description, figures, and claims are not meant to be limiting. Other embodiments can be utilized, and other changes can be made, without departing from the spirit or scope of the subject matter presented herein. It will be readily understood that the aspects of the present disclosure, as generally described herein, and illustrated in the figures, can be arranged, substituted, combined, separated, and designed in a wide variety of different configurations, all of which are explicitly contemplated herein.

[0157] Numerous modifications and variations of the present disclosure are possible in light of the above teachings. Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements. Groupings of alternative elements or embodiments of the invention disclosed herein are not to be construed as limitations. Each group member may be referred to and claimed individually or in any combination with other members of the group or other elements found herein. It is anticipated that one or more members of a group may be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

[0158] Furthermore, numerous references have been made to patents and printed publications throughout this specification. Each of the above-cited references and printed publications are individually incorporated herein by reference in their entirety.

[0159] It is to be understood that the embodiments of the invention disclosed herein are illustrative of the principles of the present invention. Other modifications that may be employed are within the scope of the invention. Thus, by way of example, but not of limitation, alternative configurations of the present invention may be utilized in accordance with the teachings herein.

[0160] Accordingly, the present invention is not limited to that precisely as shown and described. The particulars shown herein are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of various embodiments of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for the fundamental understanding of the invention, the description taken with the drawings and/or examples making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

REFERENCES

- [0161] 1. N. P. King et al., Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 336, 1171 (Jun. 1, 2012).
- [0162] 2. N. P. King et al., Accurate design of co-assembling multi-component protein nanomaterials. *Nature* 510, 103 (Jun. 5, 2014).
- [0163] 3. S. Raman et al., Design of Peptide Nanoparticles Using Simple Protein Oligomerization Domains. *The Open Nanomedicine Journal* 2, 15 (2009).
- [0164] 4. J. P. Julien et al., Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342, 1477 (Dec. 20, 2013).
- [0165] 5. D. Lyumkis et al., Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* 342, 1484 (Dec. 20, 2013).
- [0166] 6. M. Pancera et al., Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*, (Oct. 8, 2014).
- [0167] 7. T. O. Yeates, C. S. Crowley, S. Tanaka, Bacterial microcompartment organelles: protein shell structure and evolution. *Annu Rev Biophys* 39, 185 (2010).
- [0168] 8. P. Kumar, M. Singh, S. Karthikeyan, Crystal structure analysis of icosahedral lumazine synthase from *Salmonella typhimurium*, an antibacterial drug target. *Acta Crystallogr D Biol Crystallogr* 67, 131 (February, 2011).
- [0169] 9. C. Jäckel, J. D. Bloom, P. Kast, F. H. Arnold, D. Hilvert. Consensus Protein Design without Phylogenetic Bias *J. Mol. Biol.*, 399 (2010), pp. 541-546.
- [0170] 10. Hura, G. L. et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* 6, 606-612 (2009).
- [0171] 11. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys.* 1 105, 962-974 (2013).
- [0172] 12. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* 38, W540-W544 (2010).

SEQUENCE LISTING

Sequence total quantity: 43

SEQ ID NO: 1 moltype = AA length = 207
 FEATURE Location/Qualifiers
 REGION 1..207
 note = Synthetic
 source 1..207
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 1
 MEGMDPLAVL AESRLPLLT VRGGEDLAGL ATVLELMGVG ALEITLRTEK GLEALKALRK 60
 SGLLLGAGTV RSPKEAEAAL EAGAFLVSP GLLEEVAALA QARGVPYLPV VLTPTEVERA 120
 LALGLSALKF FPAEPFQGVV VLRAYAEVFP EVRFLPTGGI KEEHLPHYAA LPNLLAVGGS 180
 WLLQGDAAV MKKVKAAL LSPQAPG 207

SEQ ID NO: 2 moltype = AA length = 156
 FEATURE Location/Qualifiers
 REGION 1..156
 note = Synthetic
 source 1..156
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 2
 MTKKVGIVDT TFARVMAEA AIRTLKALSP NIKIIRKTPV GIKDLPVACK KLEEEGCDI 60
 VMALGMPGKA EKDKVCAEA SLGLMLAQLM TNKHIEVFV HEDEAKDDDE LDILALVRAI 120
 EHAANVYYLL FKPEYLTRMA GKGLRQGRED AGPARE 156

SEQ ID NO: 3 moltype = AA length = 156
 FEATURE Location/Qualifiers
 REGION 1..156
 note = Synthetic
 source 1..156
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 3

-continued

MTKKVGIVDT	TFARVDMASA	AILTLKMESP	NIKIIRKTV	GIKDLPVACK	KLLEEEGCDI	60
VMALGMPGKA	EKDKVCAHEA	SLGLMLAQLM	TNKHIEVFV	HEDEAKDDAE	LKILAARRAI	120
EHALNVYYLL	FKPEYLTRMA	GKGLRQGFED	AGPARE			156
SEQ ID NO: 4						
FEATURE						
REGION						
source						
SEQUENCE: 4						
MSTINNQLKA	LKVIPVIAID	NAEDIIPLGK	VLAENGLPAA	EITFRSSAAV	KAIMLLRSAQ	60
PEMLIGAGTI	LNGVQALAAK	EAGATFVVSP	GFNPNTVRAC	QIIGIDIVPG	VMNPSTVEAA	120
LEMGLTTLKF	FPAEASGGIS	MVKSLVGPYG	DIRLMPTGGI	TPSNIDNYLA	IPQVLACGGT	180
WMVDKKLVTN	GEWDEIARLT	REIVEQVNP				209
SEQ ID NO: 5						
FEATURE						
REGION						
source						
SEQUENCE: 5						
MPIFTLNTNI	KATDVPSDFL	SLTSRLVGLI	LSKPGSYVAV	HINTDQQLSF	GGSTNPAAFG	60
TLMSIGGIEP	SKNRDHSAVL	FDHLNAMLGI	PKNRMYIHFV	NLNGDDVGWN	GTTF	114
SEQ ID NO: 6						
FEATURE						
REGION						
source						
SEQUENCE: 6						
MNQSHSKDYE	TVRIAVVRAR	WHADIVDACV	EAFEIAMA	GGDRFAVDVF	DVPGAYE IPL	60
HARTLAETGR	YGAVLGTA	VNGGIYRHEF	VASAVIDGMM	NVQLSTGVPV	LSAVLTPHRY	120
RDSA EHRFF	AAHFAVKGVE	AARACIEILA	AREKIAA			157
SEQ ID NO: 7						
FEATURE						
REGION						
source						
SEQUENCE: 7						
MKMEELFKKH	KIVAVLRANS	VEEAIEKAVA	VFAGGVHLIE	ITFTVPDADT	VIKALSVLKE	60
KGAIIGAGTV	TSVEQCRKAV	ESGAEFIVSP	HLDEEISQFC	KEKGVFYMPG	VMTPTLVKA	120
MKLGHTILKL	FPGEVGPQF	VKAMKGPFPN	VKFVPTGGVN	LDNVCEWFKA	GVLA VGVGSA	180
LVKGTPDEV	EKAKAFVEKI	RGCTE				205
SEQ ID NO: 8						
FEATURE						
REGION						
source						
SEQUENCE: 8						
MNQSHSKDYE	TVRIAVVRAR	WHAEIVDACV	SAFEAAMADI	GGDRFAVDVF	DVPGAYE IPL	60
HARTLAETGR	YGAVLGTA	VNGGIYRHEF	VASAVIDGMM	NVQLSTGVPV	LSAVLTPHRY	120
RSDAHTLLF	LALFAVKGME	AARACVEILA	AREKIAA			157
SEQ ID NO: 9						
FEATURE						
REGION						
source						
SEQUENCE: 9						
MFTKSGDDGN	TNVINKRVGK	DSPLVNFLGD	LDELNSFIGF	AISKIPWEDM	KKDLERVQVE	60
LFEIGEDLST	QSSKKKIDES	YVLWLLAATA	IYRIESGPVK	LFVIPGGSEE	ASVLHVTRSV	120
ARRVERNAVK	YTKELPEINR	MIIVYLNRLS	SLLFAMALVA	NKRRNQSEKI	YEIGKSW	177

-continued

SEQ ID NO: 10	moltype = AA length = 157					
FEATURE	Location/Qualifiers					
REGION	1..157					
	note = Synthetic					
source	1..157					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 10						
MNQHSHKDYE	TVRIAVVRAR	WHADIVDQCV	RAFEEAMADA	GGDRFAVDVF	DVPGAYEIPL	60
HARTLAETGR	YGAVLGTA	FV VNGGIYRHEF	VASAVIDGMM	NVQLSTGVPV	LSAVLTPHRY	120
RSSREHHEFF	REHFMVKGVE	AAAACITILA	AREKIAA			157
SEQ ID NO: 11	moltype = AA length = 201					
FEATURE	Location/Qualifiers					
REGION	1..201					
	note = Synthetic					
source	1..201					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 11						
MGHTKGPTPQ	QHDGSALRIG	IVHARWNKTI	IMPLLLIGTIA	KLLECGVKAS	NIVVQSVPGS	60
WELPIAVQRL	YSASQLQTPS	SGPSLSAGDL	LGSSTTDLTA	LPTTTASSTG	PFDALIAIGV	120
LIKGETMHFE	YIADSVSHGL	MRVQLDTGVP	VIFGVLT	VLTT DDQAKARAGV	IEGSHNHGED	180
WGLAAVEMGV	RRRDWAAGKT	E				201
SEQ ID NO: 12	moltype = AA length = 237					
FEATURE	Location/Qualifiers					
REGION	1..237					
	note = Synthetic					
source	1..237					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 12						
MYEVDHADVY	DLFYLGRGKD	YAAEASDIAD	LVRSRTP	PEAS SLLDVACGTG	THLEHFTKEF	60
GDTAGLELSE	DMLTHARKRL	PDATLHQGDM	RDFQLGRKFS	AVVSMFSSVG	YLKTVAELGA	120
AVASFAEHLE	PGGVVVVEPW	WFPETFADGW	VSADVVRD	RTVARVSHSV	REGNATRMEV	180
HFTVADPGKG	VRHFSDVHLI	TLFHQREYEA	AFMAAGLRVE	YLEGGPSGRG	LFVGVPA	237
SEQ ID NO: 13	moltype = AA length = 138					
FEATURE	Location/Qualifiers					
REGION	1..138					
	note = Synthetic					
source	1..138					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 13						
MGMKEKPVLI	ITHGDFGKGL	LSGAEVIIGK	QENVHTVGLN	LGDNIEKVAK	EVMRIIIAKL	60
AEDKEIIIVV	DLFGGSPFNI	ALEMMKTFDV	KVITGINMPM	LVELLTSINV	YDTTELLENI	120
SKIGKDGIKV	IEKSSLKM					138
SEQ ID NO: 14	moltype = AA length = 154					
FEATURE	Location/Qualifiers					
REGION	1..154					
	note = Synthetic					
source	1..154					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 14						
MKYDGSKLRI	GILHARWNLE	IIAALVAGAI	KRLQEF	GVKA ENIIIIETVPG	SFELPYGSKL	60
FVEKQKRLGK	PLDAIIPIGV	LIKGSTMHFE	YICDSTTHQL	MKLN	FELGIP VIFGVLTCLT	120
DEQAEARAGL	IEGKMHNHGE	DWGAAAVEMA	TKFN			154
SEQ ID NO: 15	moltype = AA length = 164					
FEATURE	Location/Qualifiers					
REGION	1..164					
	note = Synthetic					
source	1..164					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 15						
MAVKGLGEVD	QKYDGSKLRI	GILHARWNRK	IILALVAGAV	LRLLEFGVKA	ENIIIIETVPG	60
SFELPYGSKL	FVEKQKRLGK	PLDAIIPIGV	LIKGSTMHFE	YICDSTTHQL	MKLN	FELGIP 120
VIFGVLTCLT	DEQAEARAGL	IEGKMHNHGE	DWGAAAVEMA	TKFN		164
SEQ ID NO: 16	moltype = AA length = 175					

-continued

FEATURE	Location/Qualifiers				
REGION	1..175				
	note = Synthetic				
source	1..175				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 16					
MGANWYLDNE	SSRLSFTSTK	NADIAEVHRF	LVLHGKVDPK	GLAEVEVETE	SISTGIPLRD 60
MLLRVLVFQV	SKFPVAQINA	QLDMRPINNL	APGAQLELRL	PLTVSLRGKS	HSYNAELLAT 120
RLDERRFQVV	TLEPLVIHAQ	DFDMVRAFNA	LRLVAGLSAV	SLSVPVGAVL	IFTAR 175
SEQ ID NO: 17	moltype = AA length = 208				
FEATURE	Location/Qualifiers				
REGION	1..208				
	note = Synthetic				
source	1..208				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 17					
MTDYIRDGSA	IKALSFAIL	AEADLRHIPQ	DLQRLAVRVI	HACGMVDVAN	DLAFSEGAGK 60
AGRNALLAGA	PILCDARMVA	EGITRSRLPA	DNRVIYTLSD	PSVPELAKKI	GNTRSAAALD 120
LWLPHIEGSI	VAIGNAPTAL	FRLFELLDAG	APKPALIIGM	PVGFGAAES	KDELAANSRG 180
VPYVIVRGRR	GGSAMTAAAV	NALASERE			208
SEQ ID NO: 18	moltype = AA length = 128				
FEATURE	Location/Qualifiers				
REGION	1..128				
	note = Synthetic				
source	1..128				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 18					
MITVFGLKSK	LAPPREKLAE	VIYSSLHLGL	DIPKGKHAIR	FLCLEKEDFY	YPFDRSDDYT 60
VIEINLMAGR	SEETKMLLIF	LLFIALERKL	GIRAHDVEIT	IKEQPAHCWG	FRGRTGDSAR 120
DLDYDIYV					128
SEQ ID NO: 19	moltype = AA length = 235				
FEATURE	Location/Qualifiers				
REGION	1..235				
	note = Synthetic				
source	1..235				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 19					
MGSDLQKLQR	FSTCDISDGL	LNVYNIPTGG	YFPNLTAISP	PQNSSIVGTA	YTVLFAPIDD 60
PRPAVNYIDS	VPPNSILVLA	LEPHLQSQFH	PFIKITQAMY	GGLMSTRAQY	LKSNGTVVFG 120
RIRDVDEHRT	LNHPVFAYGV	GSCAPKAVVK	AVGTNVQLKI	LTSDGVTQTI	CPGDYIAGDN 180
NGIVRIPVQE	TDISKLVTYI	EKSIEVDRLV	SEAIKNGLPA	KAAQTARRMV	LKDYI 235
SEQ ID NO: 20	moltype = AA length = 162				
FEATURE	Location/Qualifiers				
REGION	1..162				
	note = Synthetic				
source	1..162				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 20					
MSGMRVYLGA	DHAGYELKQA	IIAFLKMTGH	EPIDCGALRY	DADDDYP AFC	IAAATRTVAD 60
PGSLGIVLGG	SGNGEQIAAN	KVPGARCALA	WSVQTAALAR	EHNNAQLIGI	GGRMHTLEEA 120
LRIVKAFVTT	PWSKAQRHQR	RIDILAEYER	THEAPPVPGA	PA	162
SEQ ID NO: 21	moltype = AA length = 157				
FEATURE	Location/Qualifiers				
REGION	1..157				
	note = Synthetic				
source	1..157				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 21					
MGDDARIAAI	GDVDELNSQI	GVL LAEPLPD	DVRAALSAIQ	HDLFDLG GEL	CIPGHAAITE 60
DHLLRLALWL	VHYNGQLPPL	EEFILPGGAR	GAALAHVCRT	VCRR AERSIK	ALGASEPLNI 120
APAAYVNLLS	DLLFVLARVL	NRAAGGADVL	WDRTRAH		157
SEQ ID NO: 22	moltype = AA length = 157				
FEATURE	Location/Qualifiers				
REGION	1..157				

-continued

source		note = Synthetic	
		1..157	
		mol_type = protein	
		organism = synthetic construct	
SEQUENCE: 22			
MILSAEQSFT	LRHPHGQAAA	LAFVREPAAA LAGVQRLRGL DSDGEQVWGE LLVRVPLLGE	60
VDLPFRSEIV	RTPQGAELRP	LTLTGERAWV AVSGQATAAE GGEMAFAPQF QAHLATPEAE	120
GEGGAAFEVM	VQAAAGVTLL	LVAMALPQGL AAGLPPA	157
SEQ ID NO: 23		moltype = AA length = 156	
FEATURE		Location/Qualifiers	
REGION		1..156	
		note = Synthetic	
source		1..156	
		mol_type = protein	
		organism = synthetic construct	
SEQUENCE: 23			
MTKKVGIVDT	TFARVDMASA	AILTLKMESP NIKIIRKTVP GIKDLPVACK KLEEEEGCDI	60
VMALGMPGKK	EKDKVCAHEA	SLGLMLAQLM TNKHIIEVFV HEDEAKDDAE LKILAARRAI	120
EHALNVYYLL	FKPEYLTRMA	GKGLRQGFED AGPARE	156
SEQ ID NO: 24		moltype = AA length = 209	
FEATURE		Location/Qualifiers	
REGION		1..209	
		note = Synthetic	
source		1..209	
		mol_type = protein	
		organism = synthetic construct	
SEQUENCE: 24			
MDDINNQLKR	LKVIPVIAID	NAEDIIPLGK VLAENGLPAA EITFRSSAAV KAIMLLRSAQ	60
PEMLIGAGTI	LNGVQALAAK	EAGADFFVSP GFNPNTVRAC QIIGIDIVPG VMNPSTVEQA	120
LEMGLTTLKF	FPAEASGGIS	MVKSLVGPYG DIRLMPTGGI TPDNIDNYLA IPQVLACGGT	180
WMVDKKLVRN	GEWDEIARLT	REIVEQVNP	209
SEQ ID NO: 25		moltype = AA length = 114	
FEATURE		Location/Qualifiers	
REGION		1..114	
		note = Synthetic	
source		1..114	
		mol_type = protein	
		organism = synthetic construct	
SEQUENCE: 25			
MPIFTLNTNI	KADDVPSDFL	SLTSRLVGLI LSKPGSYVAV HINTDQQLSF GGSTNPAAFG	60
TLMSIGGIEP	DKNRDHSAVL	FDHLNAMLGI PKNRM YIHFV NLNGDDVGWN GTTF	114
SEQ ID NO: 26		moltype = AA length = 114	
FEATURE		Location/Qualifiers	
REGION		1..114	
		note = Synthetic	
source		1..114	
		mol_type = protein	
		organism = synthetic construct	
SEQUENCE: 26			
MPIFTLNTNI	KADDVPSDFL	SLTSRLVGLI LSEPGSYVAV HINTDQQLSF GGSTNPAAFG	60
TLMSIGGIEP	DKNEDHSAVL	FDHLNAMLGI PKNRM YIHFV DLDGDDVGWN GTTF	114
SEQ ID NO: 27		moltype = AA length = 157	
FEATURE		Location/Qualifiers	
REGION		1..157	
		note = Synthetic	
source		1..157	
		mol_type = protein	
		organism = synthetic construct	
SEQUENCE: 27			
MNQSHSKDHE	TVRIAVVRAR	WHADIVDACV EAFEIAMA AI GGDRFAVDVF DVP GAYEIPL	60
HARTLAETGR	YGAVLGTA FV	VNGGIYRHEF VASAVIDGMM NVQLDTGVPV LSAVLTPHRY	120
RDSDEHHRFF	AAHFAVKGVE	AARACIEILN AREKIAA	157
SEQ ID NO: 28		moltype = AA length = 157	
FEATURE		Location/Qualifiers	
REGION		1..157	
		note = Synthetic	
source		1..157	
		mol_type = protein	
		organism = synthetic construct	
SEQUENCE: 28			

-continued

MNQSHSKDHE	TVRIAVVRAR	WHADIVDACV	EAFEIAMAAI	GGDRFAVDVF	DVPGAYE IPL	60
HARTLAETGR	YGAVLGTA FV	VDGGIYDHEF	VASAVIDGMM	NVQLDTGVPV	LSAVLTPHEY	120
EDSDEDHEFF	AAHFAVKGVE	AARACIEILN	AREKIAA			157
SEQ ID NO: 29	moltype = AA length = 205					
FEATURE	Location/Qualifiers					
REGION	1..205					
	note = Synthetic					
source	1..205					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 29						
MKMEELFKKH	KIVAVLRANS	VEEAIEKAVA	VFAGGVHLIE	ITFTVPDADT	VIKALSVLKE	60
KGAIIGAGTV	TSVEQCRKAV	ESGAEFIVSP	HLDEEISQFC	KEKGVFYMPG	VMTPTLVKA	120
MKLGH DILKL	FPGEVGPQF	VKAMKGPPFN	VK FVPTGGVN	LDNVCEWFKA	GV LAVGVGDA	180
LVKGDPDEV R	EKAKKFVEKI	RGCTE				205
SEQ ID NO: 30	moltype = AA length = 205					
FEATURE	Location/Qualifiers					
REGION	1..205					
	note = Synthetic					
source	1..205					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 30						
MKMEELFKKH	KIVAVLRANS	VEEAIEKAVA	VFAGGVHLIE	ITFTVPDADT	VIKALSVLKE	60
KGAIIGAGTV	TSVEQCRKAV	ESGAEFIVSP	HLDEEISQFC	KEKGVFYMPG	VMTPTLVKA	120
MKLGH DILKL	FPGEVGPPEF	VEAMKGPPFN	VK FVPTGGVD	LDDVCEWFDA	GV LAVGVGDA	180
LVEGDPDEV R	EDAKEFVEEI	RGCTE				205
SEQ ID NO: 31	moltype = AA length = 205					
FEATURE	Location/Qualifiers					
REGION	1..205					
	note = Synthetic					
source	1..205					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 31						
MKMEELFKKH	KIVAVLRANS	VEEAIEKAVA	VFAGGVHLIE	ITFTVPDADT	VIKALSVLKE	60
KGAIIGAGTV	TSVEQCRKAV	ESGAEFIVSP	HLDEEISQFC	KEKGVFYMPG	VMTPTLVKA	120
MKLGH DILKL	FPGEVGPQF	VKAMKGPPFN	VK FVPTGGVN	LDNVCKWFKA	GV LAVGVGKA	180
LVKGKPDEV R	EKAKKFVKKI	RGCTE				205
SEQ ID NO: 32	moltype = AA length = 157					
FEATURE	Location/Qualifiers					
REGION	1..157					
	note = Synthetic					
source	1..157					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 32						
MNQSHSKDHE	TVRIAVVRAR	WHAEIVDACV	SAFEAAMRDI	GGDRFAVDVF	DVPGAYE IPL	60
HARTLAETGR	YGAVLGTA FV	VNGGIYRHEF	VASAVIDGMM	NVQLDTGVPV	LSAVLTPHRY	120
RDSAHTLLF	LALFAVKGME	AARACVEILA	AREKIAA			157
SEQ ID NO: 33	moltype = AA length = 157					
FEATURE	Location/Qualifiers					
REGION	1..157					
	note = Synthetic					
source	1..157					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 33						
MNQSHSKDHE	TVRIAVVRAR	WHAEIVDACV	SAFEAAMRDI	GGDRFAVDVF	DVPGAYE IPL	60
HARTLAETGR	YGAVLGTA FV	VDGGIYDHEF	VASAVIDGMM	NVQLDTGVPV	LSAVLTPHEY	120
EDSDADTLLF	LALFAVKGME	AARACVEILA	AREKIAA			157
SEQ ID NO: 34	moltype = AA length = 157					
FEATURE	Location/Qualifiers					
REGION	1..157					
	note = Synthetic					
source	1..157					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 34						
MNQSHSKDHE	TVRIAVVRAR	WHAEIVDACV	SAFEAAMRDI	GGDRFAVDVF	DVPGAYE IPL	60

-continued

HARTLAETGR	YGAVLGTA	FV	VNGGIYRHEF	VASAVINGMM	NVQLNTGVPV	LSAVLTPHNY	120
DKSKAHTLLF	LALFAVKGME		AARACVEILA	AREKIAA			157
SEQ ID NO: 35	moltype = AA length = 156						
FEATURE	Location/Qualifiers						
REGION	1..156						
	note = Synthetic						
VARIANT	70						
	note = Xaa is A or K						
source	1..156						
	mol_type = protein						
	organism = synthetic construct						
SEQUENCE: 35							
MTKKVGIVDT	TFARVDMASA		AILTLKMESP	NIKIIRKTV	GIKDLPVACK	KLLEEEGCDI	60
VMALGMPGKX	EKDKVCAHEA		SLGLMLAQLM	TNKHIIEVFV	HEDEAKDDAE	LKILAARRAI	120
EHALNVYYLL	FKPEYLTRMA		GKGLRQGFED	AGPARE			156
SEQ ID NO: 36	moltype = AA length = 209						
FEATURE	Location/Qualifiers						
REGION	1..209						
	note = Synthetic						
VARIANT	2						
	note = Xaa is S or D						
VARIANT	3						
	note = Xaa is T or D						
VARIANT	10						
	note = Xaa is A or R						
VARIANT	85						
	note = Xaa is T or D						
VARIANT	119						
	note = Xaa is A or Q						
VARIANT	163						
	note = Xaa is S or D						
VARIANT	189						
	note = Xaa is T or R						
source	1..209						
	mol_type = protein						
	organism = synthetic construct						
SEQUENCE: 36							
MXXINNQLKX	LKVIPVIAID		NAEDIIPLGK	VLAENGLPAA	EITFRSSAAV	KAIMLLRSAQ	60
PEMLIGAGTI	LNGVQALAAK		EAGAXFVVSP	GFNPNTVRAC	QIIGIDIVPG	VMNPSTVEXA	120
LEMGLTTLKF	FPAEASGGIS		MVKSLVGPYG	DIRLMPTGGI	TPXNIDNYLA	IPQVLACGGT	180
WMVDDKKLVXN	GEWDEIARLT		REIVEQVNP				209
SEQ ID NO: 37	moltype = AA length = 114						
FEATURE	Location/Qualifiers						
REGION	1..114						
	note = Synthetic						
VARIANT	13						
	note = Xaa is T or D						
VARIANT	33						
	note = Xaa is K or E						
VARIANT	71						
	note = Xaa is S or D						
VARIANT	74						
	note = Xaa is R or E						
VARIANT	101						
	note = Xaa is N or D						
VARIANT	103						
	note = Xaa is N or D						
source	1..114						
	mol_type = protein						
	organism = synthetic construct						
SEQUENCE: 37							
MPIFTLNTNI	KAXDVPSDFL		SLTSRLVGLI	LSXPGSYVAV	HINTDQQLSF	GGSTNPAAFG	60
TLMSIGGIEP	XKNXDHSAVL		FDHLNAMLGI	PKNRMYIHV	XLXGDDVGWN	GTTF	114
SEQ ID NO: 38	moltype = AA length = 157						
FEATURE	Location/Qualifiers						
REGION	1..157						
	note = Synthetic						
VARIANT	9						
	note = Xaa is Y or H						
VARIANT	82						
	note = Xaa is N or D						
VARIANT	87						

-continued

	note = Xaa is R or D	
VARIANT	105	
	note = Xaa is S or D	
VARIANT	119	
	note = Xaa is R or E	
VARIANT	121	
	note = Xaa is R or E	
VARIANT	124	
	note = Xaa is A or D	
VARIANT	126	
	note = Xaa is H or D	
VARIANT	128	
	note = Xaa is R or E	
VARIANT	150	
	note = Xaa is A or N	
source	1..157	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 38		
MNQSHKDXE	TVRIAVVRAR WHADIVDACV EAFEIAMA	AI GGDRFAVDVF DVPGAYEIP
		L 60
HARTLAETGR	YGAVLGTA	FV VXGGIYXHEF VASAVIDGMM NVQLXTGVPV LSAVLTPHXY 120
XDSXEXHXFF	AAHFAVKGVE AARACIEILX AREKIAA	157
SEQ ID NO: 39	moltype = AA	length = 205
FEATURE	Location/Qualifiers	
REGION	1..205	
	note = Synthetic	
VARIANT	126	
	note = Xaa is T or D	
VARIANT	139	
	note = Xaa is Q or E	
VARIANT	142	
	note = Xaa is K or E	
VARIANT	160	
	note = Xaa is N or D	
VARIANT	163	
	note = Xaa is N or D	
VARIANT	166	
	note = Xaa is E or K	
VARIANT	169	
	note = Xaa is D or K	
VARIANT	179	
	note = Xaa is S, D or K	
VARIANT	183	
	note = Xaa is K or E	
VARIANT	185	
	note = Xaa is T, D, or K	
VARIANT	192	
	note = Xaa is D or K	
VARIANT	195	
	note = Xaa is A, E, or K	
VARIANT	198	
	note = Xaa is E or K	
VARIANT	199	
	note = Xaa is E or K	
source	1..205	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 39		
MKMEELFKKH	KIVAVLRANS VEEAIEKAVA VFAGGVHLIE ITFTVPDADT VIKALSVLKE	60
KGAIIGAGTV	TSVEQCRKAV ESGAEFIVSP HLDEEISQFC KEKGVFYMPG VMTPTELVKA	120
MMKLGHXILKL	FPGEVVGPF	X VXAMKGPFPN VKFVPTGGVX LDXVCXWFXA GVLAVGVGXA 180
LVXGXPDEV	R EXAKXFVXXI RGCTE	205
SEQ ID NO: 40	moltype = AA	length = 157
FEATURE	Location/Qualifiers	
REGION	1..157	
	note = Synthetic	
VARIANT	9	
	note = Xaa is Y or H	
VARIANT	38	
	note = Xaa is A or R	
VARIANT	82	
	note = Xaa is N or D	
VARIANT	87	
	note = Xaa is R or D	

-continued

VARIANT	97	
	note = Xaa is N or D	
VARIANT	105	
	note = Xaa is S, N, or D	
VARIANT	119	
	note = Xaa is R, E, or N	
VARIANT	121	
	note = Xaa is R, E, or D	
VARIANT	122	
	note = Xaa is K or D	
VARIANT	124	
	note = Xaa is K or D	
VARIANT	126	
	note = Xaa is H or D	
source	1..157	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 40		
MNQSHKDXE	TVRIAVVRAR WHAEIVDACV SAFEAAMXDI GGDRFAVDVF DVPGAYE IPL	60
HARTLAETGR	YGAVLGTA FV VXGGIYXHEF VASAVIXGMM NVQLXTGVPV LSAVLTPHXY	120
XXSXAXTLLF	LALFAVKGME AARACVEILA AREKIAA	157
SEQ ID NO: 41	moltype = DNA length = 142	
FEATURE	Location/Qualifiers	
misc_feature	1..142	
	note = Synthetic	
source	1..142	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 41		
taatgcttaa	gtcgaacaga aagtaatcgt attgtacacg gccgcataat cgaaattaat	60
acgactcact	ataggggaat tgtgagcgga taacaattcc ccatcttagt atattagtta	120
agtataagaa	ggagatatac tt	142
SEQ ID NO: 42	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
misc_feature	1..18	
	note = Synthetic	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 42		
taaagaagga	gatatcat	18
SEQ ID NO: 43	moltype = DNA length = 17	
FEATURE	Location/Qualifiers	
misc_feature	1..17	
	note = Synthetic	
source	1..17	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 43		
tgagaaggag	atatcat	17

- 1.-20.** (canceled)

21. A self-assembling icosahedral protein nanostructure, comprising:

20 trimeric components having 3-fold symmetry, each trimeric component comprising 3 first subunits, wherein each first subunit comprises a polypeptide having a polypeptide sequence that is at least 80% identical to SEQ ID NO: 7 or 29-31; and

12 pentameric components having 5-fold symmetry, each pentameric component comprising 5 second subunits, wherein each second subunit comprises a polypeptide having a polypeptide sequence that is at least 80% identical to SEQ ID NO: 8 or 32-34.

22. The nanostructure of claim 21, wherein the polypeptide sequence of each first subunit is at least 90% identical to SEQ ID NO: 7 or 29-31; and

wherein the polypeptide sequence of each second subunit is at least 90% identical to SEQ ID NO: 8 or 32-34.

23. The nanostructure of claim 21, wherein the polypeptide sequence of each first subunit is at least 95% identical to SEQ ID NO: 7 or 29-31; and

wherein the polypeptide sequence of each second subunit is at least 95% identical to SEQ ID NO: 8 or 32-34.

24. The nanostructure of claim 23, wherein the trimeric components and the pentameric components form an I53 icosahedral architecture.

25. The nanostructure of claim 21, wherein the polypeptide sequence of each first subunit is SEQ ID NO: 7; and

wherein the polypeptide sequence of each second subunit is SEQ ID NO: 8.

26. The nanostructure of claim 21, wherein the polypeptide sequence of each first subunit is SEQ ID NO: 7; and

