



- (51) **International Patent Classification:** Not classified
- (21) **International Application Number:**
PCT/US2016/067295
- (22) **International Filing Date:**
16 December 2016 (16.12.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/268,320 16 December 2015 (16.12.2015) US
- (71) **Applicant:** UNIVERSITY OF WASHINGTON
[US/US]; 4545 Roosevelt Way NE, Suite 400, Seattle, WA 98105 (US).
- (72) **Inventors:** PARMEGGIANI, Fabio; c/o University of Washington, 4545 Roosevelt Way NE, Suite 400, Seattle, WA 98105 (US). BRUNETTE, TJ; c/o University of Washington, 4545 Roosevelt Way NE, Suite 400, Seattle, WA 98105 (US). HUANG, Po-Ssu; c/o University of Washington, 4545 Roosevelt Way NE, Suite 400, Seattle, WA 98105 (US). BAKER, David; c/o University of Washington, 4545 Roosevelt Way NE, Suite 400, Seattle, WA 98105 (US).
- (74) **Agent:** LOOS, Thomas, J.; McDonnell Boehnen Hulbert & Berghoff LLP, 300 South Wacker Drive, Chicago, IL 60606 (US).

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

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

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** REPEAT PROTEIN ARCHITECTURES

(57) **Abstract:** Methods and systems for designing proteins are disclosed, as well as proteins and protein assemblies designed. A comparing device can determine a protein repeating unit that includes one or more protein helices and one or more protein loops. The computing device can generate a protein backbone structure with a copy of the protein repeating unit. The computing device can determine whether a distance between a pair of helices of the protein backbone structure is between lower and upper distance thresholds. After determining that the distance between the pair of helices is between the lower and upper distance thresholds, the computing device can generate a plurality of protein sequences based on the protein backbone structure, select a particular protein sequence of the plurality of protein sequences based on an energy landscape that has information about energy and distance from a target fold of the particular protein sequence, and generate an output based on the particular protein sequence.



Repeat Protein Architectures

5

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

This invention was made with U.S. government support under MCB-1445201 and CHE-1332907, awarded by National Science Foundation, under N00024-10-D-6318/0024, awarded by the Defense Threat Reduction Agency, and under FA950-12-10112, awarded by the Air Force Office of Scientific Research. The U.S. Government has certain rights in the invention.

BACKGROUND

A central question in protein evolution is the extent to which naturally occurring proteins sample the space of folded structures accessible to the polypeptide chain. Repeat proteins composed of multiple tandem copies of a modular structure unit¹ are widespread in nature and play critical roles in molecular recognition, signaling, and other essential biological processes². Naturally occurring repeat proteins have been reengineered for molecular recognition and modular scaffolding applications³⁻⁵.

20

SUMMARY OF THE INVENTION

Here we use computational protein design to investigate the space of folded structures that can be generated by tandem repeating a simple helix-loop-helix-loop structural motif. 83 designs with sequences unrelated to known repeat proteins were experimentally characterized. 53 were monomeric and stable at 95 °C, and 43 have solution x-ray scattering spectra closely consistent with the design models. Crystal structures of 15 designs spanning a broad range of curvatures are in close agreement with the design models with RMSDs ranging from 0.7 to 2.5 Å. Our results show that existing repeat proteins occupy only a small fraction of the possible repeat protein sequence and structure space and that it is possible to design novel repeat proteins with precisely specified geometries, opening up a wide array of new possibilities for biomolecular engineering.

30

In one aspect, the present invention provides polypeptides comprising or consisting of the amino acid sequence selected from the group consisting of the following multi-domain proteins, as further defined in the detailed description:

- (a) SEQ ID NO:1-[SEQ ID NO:2]_(0 or 2-19)-SEQ ID NO:3;
- 5 (b) SEQ ID NO:7-[SEQ ID NO:8]_(0 or 2-19)-SEQ ID NO:9;
- (c) SEQ ID NO:13-[SEQ ID NO:14]_(0 or 2-19)-SEQ ID NO:15;
- (d) SEQ ID NO:19-[SEQ ID NO:20]_(0 or 2-19)-SEQ ID NO:21;
- (e) SEQ ID NO:25-[SEQ ID NO:26]_(0 or 2-19)-SEQ ID NO:27;
- (f) SEQ ID NO:31-[SEQ ID NO:32]_(0 or 2-19)-SEQ ID NO:33;
- 10 (g) SEQ ID NO:37-[SEQ ID NO:38]_(0 or 2-19)-SEQ ID NO:39;
- (h) SEQ ID NO:43-[SEQ ID NO:44]_(0 or 2-19)-SEQ ID NO:45;
- (i) SEQ ID NO:49-[SEQ ID NO:50]_(0 or 2-19)-SEQ ID NO:51;
- (j) SEQ ID NO:55-[SEQ ID NO:56]_(0 or 2-19)-SEQ ID NO:57;
- (k) SEQ ID NO:61-[SEQ ID NO:62]_(0 or 2-19)-SEQ ID NO:63;
- 15 (l) SEQ ID NO:67-[SEQ ID NO:68]_(0 or 2-19)-SEQ ID NO:69;
- (m) SEQ ID NO:73-[SEQ ID NO:74]_(0 or 2-19)-SEQ ID NO:75;
- (n) SEQ ID NO:79-[SEQ ID NO:80]_(0 or 2-19)-SEQ ID NO:81;
- (o) SEQ ID NO:85-[SEQ ID NO:86]_(0 or 2-19)-SEQ ID NO:87;
- (p) SEQ ID NO:91-[SEQ ID NO:92]_(0 or 2-19)-SEQ ID NO:93;
- 20 (q) SEQ ID NO:97-[SEQ ID NO:98]_(0 or 2-19)-SEQ ID NO:99;
- (r) SEQ ID NO:103-[SEQ ID NO:104]_(0 or 2-19)-SEQ ID NO:105;
- (s) SEQ ID NO:109-[SEQ ID NO:110]_(0 or 2-19)-SEQ ID NO:111;
- (t) SEQ ID NO:115-[SEQ ID NO:116]_(0 or 2-19)-SEQ ID NO:117;
- (u) SEQ ID NO:121-[SEQ ID NO:122]_(0 or 2-19)-SEQ ID NO:123;
- 25 (v) SEQ ID NO:127-[SEQ ID NO:128]_(0 or 2-19)-SEQ ID NO:129;
- (w) SEQ ID NO:133-[SEQ ID NO:134]_(0 or 2-19)-SEQ ID NO:135;
- (x) SEQ ID NO:139-[SEQ ID NO:140]_(0 or 2-19)-SEQ ID NO:141;
- (y) SEQ ID NO:145-[SEQ ID NO:146]_(0 or 2-19)-SEQ ID NO:147;
- (z) SEQ ID NO:151-[SEQ ID NO:152]_(0 or 2-19)-SEQ ID NO:153;
- 30 (aa) SEQ ID NO:157-[SEQ ID NO:158]_(0 or 2-19)-SEQ ID NO:159;
- (bb) SEQ ID NO:163-[SEQ ID NO:164]_(0 or 2-19)-SEQ ID NO:165;
- (cc) SEQ ID NO:169-[SEQ ID NO:170]_(0 or 2-19)-SEQ ID NO:171;
- (dd) SEQ ID NO:175-[SEQ ID NO:176]_(0 or 2-19)-SEQ ID NO:177;
- (ee) SEQ ID NO:181-[SEQ ID NO:182]_(0 or 2-19)-SEQ ID NO:183;

- (ff) SEQ ID NO:187-[SEQ ID NO:188]_(0 or 2-19)-SEQ ID NO:189;
- (gg) SEQ ID NO:193-[SEQ ID NO:194]_(0 or 2-19)-SEQ ID NO:195;
- (hh) SEQ ID NO:199-[SEQ ID NO:200]_(0 or 2-19)-SEQ ID NO:201;
- (ii) SEQ ID NO:205-[SEQ ID NO:206]_(0 or 2-19)-SEQ ID NO:207;
- 5 (jj) SEQ ID NO:211-[SEQ ID NO:212]_(0 or 2-19)-SEQ ID NO:213;
- (kk) SEQ ID NO:217-[SEQ ID NO:218]_(0 or 2-19)-SEQ ID NO:219;
- (ll) SEQ ID NO:223-[SEQ ID NO:224]_(0 or 2-19)-SEQ ID NO:225;
- (mm) SEQ ID NO:229-[SEQ ID NO:230]_(0 or 2-19)-SEQ ID NO:231;
- (nn) SEQ ID NO:235-[SEQ ID NO:236]_(0 or 2-19)-SEQ ID NO:237;
- 10 (oo) SEQ ID NO:241-[SEQ ID NO:242]_(0 or 2-19)-SEQ ID NO:243;
- (pp) SEQ ID NO:247-[SEQ ID NO:248]_(0 or 2-19)-SEQ ID NO:249;
- (qq) SEQ ID NO:253-[SEQ ID NO:254]_(0 or 2-19)-SEQ ID NO:255;
- (rr) SEQ ID NO:259-[SEQ ID NO:260]_(0 or 2-19)-SEQ ID NO:261;
- (ss) SEQ ID NO:265-[SEQ ID NO:266]_(0 or 2-19)-SEQ ID NO:267;
- 15 (tt) SEQ ID NO:271-[SEQ ID NO:272]_(0 or 2-19)-SEQ ID NO:273;
- (uu) SEQ ID NO:277-[SEQ ID NO:278]_(0 or 2-19)-SEQ ID NO:278;
- (vv) SEQ ID NO:283-[SEQ ID NO:284]_(0 or 2-19)-SEQ ID NO:285;
- (ww) SEQ ID NO:289-[SEQ ID NO:290]_(0 or 2-19)-SEQ ID NO:291;
- (xx) SEQ ID NO:295-[SEQ ID NO:296]_(0 or 2-19)-SEQ ID NO:297;
- 20 (yy) SEQ ID NO:301-[SEQ ID NO:302]_(0 or 2-19)-SEQ ID NO:303;
- (zz) SEQ ID NO:307-[SEQ ID NO:308]_(0 or 2-19)-SEQ ID NO:309;
- (aaa) SEQ ID NO:313-[SEQ ID NO:314]_(0 or 2-19)-SEQ ID NO:315;
- (bbb) SEQ ID NO:319-[SEQ ID NO:320]_(0 or 2-19)-SEQ ID NO:321;
- (ccc) SEQ ID NO:325-[SEQ ID NO:326]_(0 or 2-19)-SEQ ID NO:327;
- 25 (ddd) SEQ ID NO:331-[SEQ ID NO:332]_(0 or 2-19)-SEQ ID NO:333;
- (eee) SEQ ID NO:337-[SEQ ID NO:338]_(0 or 2-19)-SEQ ID NO:339;
- (fff) SEQ ID NO:343-[SEQ ID NO:344]_(0 or 2-19)-SEQ ID NO:345;
- (ggg) SEQ ID NO:349-[SEQ ID NO:350]_(0 or 2-19)-SEQ ID NO:351;
- (hhh) SEQ ID NO:355-[SEQ ID NO:356]_(0 or 2-19)-SEQ ID NO:357;
- 30 (iii) SEQ ID NO:361-[SEQ ID NO:362]_(0 or 2-19)-SEQ ID NO:363;
- (jjj) SEQ ID NO:367-[SEQ ID NO:368]_(0 or 2-19)-SEQ ID NO:369;
- (kkk) SEQ ID NO:373-[SEQ ID NO:374]_(0 or 2-19)-SEQ ID NO:375;
- (lll) SEQ ID NO:379-[SEQ ID NO:380]_(0 or 2-19)-SEQ ID NO:381;
- (mmm) SEQ ID NO:385-[SEQ ID NO:386]_(0 or 2-19)-SEQ ID NO:387;

(nnn) SEQ ID NO:391-[SEQ ID NO:392]_(0 or 2-19)-SEQ ID NO:393;
 (ooo) SEQ ID NO:397-[SEQ ID NO:398]_(0 or 2-19)-SEQ ID NO:399;
 (ppp) SEQ ID NO:403-[SEQ ID NO:404]_(0 or 2-19)-SEQ ID NO:405; and
 (qqq) SEQ ID NO:409-[SEQ ID NO:410]_(0 or 2-19)-SEQ ID NO:411;

5 wherein the domain in brackets is an optional internal domain.

In one embodiment, polypeptide comprises or consists of the amino acid sequence selected from the group consisting of:

- (A) SEQ ID NO:4-[SEQ ID NO:5]_(0 or 2-19)-SEQ ID NO:6;
- (B) SEQ ID NO:10-[SEQ ID NO:11]_(0 or 2-19)-SEQ ID NO:12;
- 10 (C) SEQ ID NO:16-[SEQ ID NO:17]_(0 or 2-19)-SEQ ID NO:18;
- (D) SEQ ID NO:22-[SEQ ID NO:23]_(0 or 2-19)-SEQ ID NO:24;
- (E) SEQ ID NO:28-[SEQ ID NO:29]_(0 or 2-19)-SEQ ID NO:30;
- (F) SEQ ID NO:34-[SEQ ID NO:35]_(0 or 2-19)-SEQ ID NO:36;
- (G) SEQ ID NO:40-[SEQ ID NO:41]_(0 or 2-19)-SEQ ID NO:42;
- 15 (H) SEQ ID NO:46-[SEQ ID NO:47]_(0 or 2-19)-SEQ ID NO:48;
- (I) SEQ ID NO:52-[SEQ ID NO:53]_(0 or 2-19)-SEQ ID NO:54;
- (J) SEQ ID NO:58-[SEQ ID NO:59]_(0 or 2-19)-SEQ ID NO:60;
- (K) SEQ ID NO:64-[SEQ ID NO:65]_(0 or 2-19)-SEQ ID NO:66;
- (L) SEQ ID NO:70-[SEQ ID NO:71]_(0 or 2-19)-SEQ ID NO:72;
- 20 (M) SEQ ID NO:76-[SEQ ID NO:77]_(0 or 2-19)-SEQ ID NO:78;
- (N) SEQ ID NO:82-[SEQ ID NO:83]_(0 or 2-19)-SEQ ID NO:84;
- (O) SEQ ID NO:88-[SEQ ID NO:89]_(0 or 2-19)-SEQ ID NO:90;
- (P) SEQ ID NO:94-[SEQ ID NO:95]_(0 or 2-19)-SEQ ID NO:96;
- (Q) SEQ ID NO:100-[SEQ ID NO:101]_(0 or 2-19)-SEQ ID NO:102;
- 25 (R) SEQ ID NO:106-[SEQ ID NO:107]_(0 or 2-19)-SEQ ID NO:108;
- (S) SEQ ID NO:112-[SEQ ID NO:113]_(0 or 2-19)-SEQ ID NO:114;
- (T) SEQ ID NO:118-[SEQ ID NO:119]_(0 or 2-19)-SEQ ID NO:120;
- (U) SEQ ID NO:124-[SEQ ID NO:125]_(0 or 2-19)-SEQ ID NO:126;
- (V) SEQ ID NO:130-[SEQ ID NO:131]_(0 or 2-19)-SEQ ID NO:132;
- 30 (W) SEQ ID NO:136-[SEQ ID NO:137]_(0 or 2-19)-SEQ ID NO:138;
- (X) SEQ ID NO:142-[SEQ ID NO:143]_(0 or 2-19)-SEQ ID NO:144;
- (Y) SEQ ID NO:148-[SEQ ID NO:149]_(0 or 2-19)-SEQ ID NO:150;
- (Z) SEQ ID NO:154-[SEQ ID NO:155]_(0 or 2-19)-SEQ ID NO:156;
- (AA) SEQ ID NO:160-[SEQ ID NO:161]_(0 or 2-19)-SEQ ID NO:162;

- (BB) SEQ ID NO:166-[SEQ ID NO:167]_(0 or 2-19)-SEQ ID NO:168;
 (CC) SEQ ID NO:172-[SEQ ID NO:173]_(0 or 2-19)-SEQ ID NO:174;
 (DD) SEQ ID NO:178-[SEQ ID NO:179]_(0 or 2-19)-SEQ ID NO:180;
 (EE) SEQ ID NO:184-[SEQ ID NO:185]_(0 or 2-19)-SEQ ID NO:186;
 5 (FF) SEQ ID NO:190-[SEQ ID NO:191]_(0 or 2-19)-SEQ ID NO:192;
 (GG) SEQ ID NO:196-[SEQ ID NO:197]_(0 or 2-19)-SEQ ID NO:198;
 (HH) SEQ ID NO:202-[SEQ ID NO:203]_(0 or 2-19)-SEQ ID NO:204;
 (II) SEQ ID NO:208-[SEQ ID NO:209]_(0 or 2-19)-SEQ ID NO:210;
 (JJ) SEQ ID NO:214-[SEQ ID NO:215]_(0 or 2-19)-SEQ ID NO:216;
 10 (KK) SEQ ID NO:220-[SEQ ID NO:221]_(0 or 2-19)-SEQ ID NO:222;
 (LL) SEQ ID NO:226-[SEQ ID NO:227]_(0 or 2-19)-SEQ ID NO:228;
 (MM) SEQ ID NO:232-[SEQ ID NO:233]_(0 or 2-19)-SEQ ID NO:234;
 (NN) SEQ ID NO:238-[SEQ ID NO:239]_(0 or 2-19)-SEQ ID NO:240;
 (OO) SEQ ID NO:244-[SEQ ID NO:245]_(0 or 2-19)-SEQ ID NO:246;
 15 (PP) SEQ ID NO:250-[SEQ ID NO:251]_(0 or 2-19)-SEQ ID NO:252;
 (QQ) SEQ ID NO:256-[SEQ ID NO:257]_(0 or 2-19)-SEQ ID NO:258;
 (RR) SEQ ID NO:262-[SEQ ID NO:263]_(0 or 2-19)-SEQ ID NO:264;
 (SS) SEQ ID NO:268-[SEQ ID NO:269]_(0 or 2-19)-SEQ ID NO:270;
 (TT) SEQ ID NO:274-[SEQ ID NO:275]_(0 or 2-19)-SEQ ID NO:276;
 20 (UU) SEQ ID NO:280-[SEQ ID NO:281]_(0 or 2-19)-SEQ ID NO:282;
 (VV) SEQ ID NO:286-[SEQ ID NO:287]_(0 or 2-19)-SEQ ID NO:288;
 (WW) SEQ ID NO:292-[SEQ ID NO:293]_(0 or 2-19)-SEQ ID NO:294;
 (XX) SEQ ID NO:298-[SEQ ID NO:299]_(0 or 2-19)-SEQ ID NO:300;
 (YY) SEQ ID NO:304-[SEQ ID NO:305]_(0 or 2-19)-SEQ ID NO:306;
 25 (ZZ) SEQ ID NO:310-[SEQ ID NO:311]_(0 or 2-19)-SEQ ID NO:312;
 (AAA) SEQ ID NO:316-[SEQ ID NO:317]_(0 or 2-19)-SEQ ID NO:318;
 (BBB) SEQ ID NO:322-[SEQ ID NO:323]_(0 or 2-19)-SEQ ID NO:324;
 (CCC) SEQ ID NO:328-[SEQ ID NO:329]_(0 or 2-19)-SEQ ID NO:330;
 (DDD) SEQ ID NO:334-[SEQ ID NO:335]_(0 or 2-19)-SEQ ID NO:336;
 30 (EEE) SEQ ID NO:340-[SEQ ID NO:341]_(0 or 2-19)-SEQ ID NO:342;
 (FFF) SEQ ID NO:346-[SEQ ID NO:347]_(0 or 2-19)-SEQ ID NO:348;
 (GGG) SEQ ID NO:352-[SEQ ID NO:353]_(0 or 2-19)-SEQ ID NO:354;
 (HHH) SEQ ID NO:358-[SEQ ID NO:359]_(0 or 2-19)-SEQ ID NO:360;
 (III) SEQ ID NO:364-[SEQ ID NO:365]_(0 or 2-19)-SEQ ID NO:366;

(JJJ) SEQ ID NO:370-[SEQ ID NO:371]_(0 or 2-19)-SEQ ID NO:372;
 (KKK) SEQ ID NO:376-[SEQ ID NO:377]_(0 or 2-19)-SEQ ID NO:378;
 (LLL) SEQ ID NO:382-[SEQ ID NO:383]_(0 or 2-19)-SEQ ID NO:384;
 (MMM) SEQ ID NO:388-[SEQ ID NO:389]_(0 or 2-19)-SEQ ID NO:390;
 5 (NNN) SEQ ID NO:394-[SEQ ID NO:395]_(0 or 2-19)-SEQ ID NO:396;
 (OOO) SEQ ID NO:400-[SEQ ID NO:401]_(0 or 2-19)-SEQ ID NO:402;
 (PPP) SEQ ID NO:406-[SEQ ID NO:407]_(0 or 2-19)-SEQ ID NO:408; and
 (QQQ) SEQ ID NO:412-[SEQ ID NO:413]_(0 or 2-19)-SEQ ID NO:414;
 wherein the domain in brackets is an optional internal domain.

10 In one embodiment, the optional internal domain may be absent. In another embodiment, the optional internal domain is present in 2-19 copies, such as in 2-3 copies.

In another aspect, the invention provides polypeptides comprising or consisting of a polypeptide having at least 50% identity over its length with the amino acid sequence selected from the group consisting of SEQ ID NO: 415-497. In various further embodiments,
 15 the polypeptides comprise or consist of a polypeptide having at least 75% identity, 90% identity, or 100% identity over its length with the amino acid sequence selected from the group consisting of SEQ ID NO: 415-497.

In another embodiment, the invention provides a protein assembly comprising a plurality of polypeptides of the invention having the same amino acid sequence. In various
 20 further embodiments, the invention provides recombinant nucleic acids encoding a polypeptides of the invention, recombinant expression vectors comprising the nucleic acid of the invention operatively linked to a promoter, and recombinant host cells comprising the recombinant expression vectors of the invention.

In one aspect, a method is provided. A computing device determines a protein
 25 repeating unit. The protein repeating unit includes one or more protein helices and one or more protein loops. The computing device generates a protein backbone structure that includes at least one copy of the protein repeating unit. The computing device determines whether a distance between a pair of helices of the protein backbone structure is between a lower distance threshold and an upper distance threshold. After determining that the
 30 distance between the pair of helices of the protein backbone structure is between the lower distance threshold and the upper distance threshold, the computing device is used for: generating a plurality of protein sequences based on the protein backbone structure, selecting a particular protein sequence of the plurality of protein sequences based on an energy landscape for the particular protein sequence, where the energy landscape includes

information about energy and distance from a target fold of the particular protein sequence, and generating an output based on the particular protein sequence.

In another aspect, a computing device is provided. The computing device includes one or more data processors and a computer-readable medium, configured to store at least
5 computer-readable instructions that, when executed, cause the computing device to perform functions. The functions include: determining a protein repeating unit, where the protein repeating unit includes one or more protein helices and one or more protein loops; generating a protein backbone structure that includes at least one copy of the protein repeating unit; determining whether a distance between a pair of helices of the protein
10 backbone structure is between a lower distance threshold and an upper distance threshold; and after determining that the distance between the pair of helices of the protein backbone structure is between the lower distance threshold and the upper distance threshold, using the computing device for: generating a plurality of protein sequences based on the protein backbone structure, selecting a particular protein sequence of the plurality of protein
15 sequences based on an energy landscape for the particular protein sequence, where the energy landscape includes information about energy and distance from a target fold of the particular protein sequence, and generating an output based on the particular protein sequence.

In another aspect, a computer-readable medium is provided. The computer-readable
20 medium is configured to store at least computer-readable instructions that, when executed by one or more processors of a computing device, cause the computing device to perform functions. The functions include: determining a protein repeating unit, where the protein repeating unit includes one or more protein helices and one or more protein loops; generating a protein backbone structure that includes at least one copy of the protein
25 repeating unit; determining whether a distance between a pair of helices of the protein backbone structure is between a lower distance threshold and an upper distance threshold; and after determining that the distance between the pair of helices of the protein backbone structure is between the lower distance threshold and the upper distance threshold, using the computing device for: generating a plurality of protein sequences based on the protein
30 backbone structure, selecting a particular protein sequence of the plurality of protein sequences based on an energy landscape for the particular protein sequence, where the energy landscape includes information about energy and distance from a target fold of the particular protein sequence, and generating an output based on the particular protein sequence.

In another aspect, a device is provided. The device comprises: means for determining a protein repeating unit, where the protein repeating unit includes one or more protein helices and one or more protein loops; means for generating a protein backbone structure that includes at least one copy of the protein repeating unit; means for determining
 5 whether a distance between a pair of helices of the protein backbone structure is between a lower distance threshold and an upper distance threshold; and means for, after determining that the distance between the pair of helices of the protein backbone structure is between the lower distance threshold and the upper distance threshold: generating a plurality of protein sequences based on the protein backbone structure, selecting a particular protein sequence
 10 of the plurality of protein sequences based on an energy landscape for the particular protein sequence, where the energy landscape includes information about energy and distance from a target fold of the particular protein sequence, and generating an output based on the particular protein sequence.

15 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Schematic overview of the computational design method. The lengths of each helix and loop were systematically enumerated. For each choice of (a) helix and loop lengths, individual repeat units (red boxes on right) were built up from fragments of proteins of known structure, and then propagated to generate extended (b) repeating structures (gray)
 20 with right-handed or left-handed twist.

Figure 2: Characterization of designed repeat proteins. (a), overall summary. Values for subset with disulfide bonds are in parantheses. (b), results on six representative designs. Top row (c): design models. Second row (d): computed energy landscapes. Energy is on y axis (REU, Rosetta energy unit) and RMSD from design model on x axis. All six landscapes
 25 are strongly funneled into the designed energy minimum. Third row (e): CD spectra collected at 25°C (red), 95°C (blue) and back to 25°C (black). The proteins do not denature within this temperature range (MRE, mean residue ellipticity; $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1} \cdot \text{residue}^{-1}$). Bottom row (f): SEC elution profile directly after affinity chromatography purification. The designs are mostly monodisperse. The maximum absorbance at 280 nm was normalized to 1.

30 Figure 3: Crystal structures of fifteen designs are in close agreement with the design models. Crystal structures are in yellow, and the design models in grey. Insets in circles show the overall shape of the repeat protein. The RMSD values across all backbone heavy atoms are: (a) 1.50 Å (DHR4), (b) 1.73 Å (DHR5), (c) 1.30 Å (DHR7), (d) 2.28 Å (DHR8), (e)

1.79 Å (DHR10), **(f)** 2.38 Å (DHR14), **(g)** 1.21 Å (DHR18), **(h)** 0.87 Å (DHR49), **(i)** 1.33 Å (DHR53), **(j)** 0.93 Å (DHR54), **(k)** 1.54 Å (DHR64), **(l)** 0.67 Å (DHR71), **(m)** 1.73 Å (DHR76), **(n)** 1.04 Å (DHR79), **(o)** 0.65 Å (DHR81). Hydrophobic side chains in the crystal structures (in red) are largely captured by the designs (Fig. 6).

5 Figure 4: Computational protocol for designing de novo repeat proteins. **(a)**, flowchart of the design protocol. The green box indicates user-controlled inputs, the grey boxes represent steps where protein structure is created or modified, and the white boxes indicate where structures are filtered. **(b)**, low resolution backbone build. **(c)**, quick full-atom design (grey) improves the backbone model (red). The superposition in the middle highlights
10 the structural changes introduced. **(d)**, structural profile: a 9-residue fragment is matched against the PDB repository for structures within 0.5 Å RMSD. The sequences from these structures are used to generate a sequence profile that influences design. **e**, packing filters were used to discard designs with cavities in the core, illustrated as grey spheres.

Figure 5: Model validation by in silico folding. To assess folding robustness seven
15 sequence variants were made for each design. **(a-g)** illustrate the energy landscape explored by Rosetta ab-initio. In red are the protein models produced by ab initio search, in green by side chain repacking and minimization (relax). Models in deep global energy minima near the relaxed structures are considered folded. The variant with highest density of ab initio models near the relax region was chosen for experimental characterization (blue box). **(h)**,
20 Jalview sequence alignment of the first 100 residues of the variants (from top to bottom: SEQ ID NOs: 581-588). The yellow bar height indicates sequence conservation, while the black bar how often the consensus sequence occurs.

Figure 6: Superposition between single internal repeats (second repeat) of designs (grey) and crystal structures (yellow). **(a)** 1.50 Å (DHR4), **(b)** 1.73 Å (DHR5), **(c)** 1.30 Å
25 (DHR7), **(d)** 2.28 Å (DHR8), **(e)** 1.79 Å (DHR10), **(f)** 2.38 Å (DHR14), **(g)** 1.21 Å (DHR18), **(h)** 0.87 Å (DHR49), **(i)** 1.33 Å (DHR53), **(j)** 0.93 Å (DHR54), **(k)** 1.54 Å (DHR64), **(l)** 0.67 Å (DHR71), **(m)** 1.73 Å (DHR76), **(n)** 1.04 Å (DHR79), **(o)** 0.65 Å (DHR81). Aliphatic and aromatic side chains are in red and cysteines are in orange. DHR7 and 18 show intra repeat disulphide bonds while DHR4 and 81 form inter-repeat cysteines.
30 DHR5 does not form the expected S-S bond. Core side chains in design recapitulate the conformation observed in the crystal structures. Even when the backbone is shifted (e.g. DHR5, 8, 15), rotamers are by large correctly predicted.

Figure 7: Designs are stable to chemical denaturation by guanidine HCl (GuHCl). Circular dichroism monitored GuHCl denaturant experiments were carried for two designs

for which crystal structures were solved (DHR4 and DHR14), two with overall shapes confirmed by SAXS (DHR21 and DHR62), and two with overall shapes inconsistent with SAXS (DHR17 and DHR67). In contrast to almost all native proteins, four of the six proteins do not denature at GuHCl concentrations up to 7.5 M. Both designs not confirmed by SAXS
5 were extremely stable to GuHCl denaturation and hence are very well folded proteins; the discrepancies between the computed and experimental SAXS profiles may be due to small amounts of oligomeric species or variation in overall twist.

Figure 8 is a block diagram of an example computing network.

Figure 9A is a block diagram of an example computing device.

10 Figure 9B depicts an example cloud-based server system.

Figure 10 is a flow chart of an example method.

DETAILED DESCRIPTION

All references cited are herein incorporated by reference in their entirety. Within this
15 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, CA), "Guide to Protein Purification" in *Methods in Enzymology* (M.P. Deutscher, ed., (1990)
20 Academic Press, Inc.); *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, CA), *Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed.* (R.I. Freshney. 1987. Liss, Inc. New York, NY), *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, TX).

25 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.

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
30 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).

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

In a first aspect, the present disclosure provides polypeptides comprising or consisting of the amino acid sequence selected from the group consisting of:

- 5 (a) SEQ ID NO:1-[SEQ ID NO:2]_(0 or 2-19)-SEQ ID NO:3;
- (b) SEQ ID NO:7-[SEQ ID NO:8]_(0 or 2-19)-SEQ ID NO:9;
- (c) SEQ ID NO:13-[SEQ ID NO:14]_(0 or 2-19)-SEQ ID NO:15;
- (d) SEQ ID NO:19-[SEQ ID NO:20]_(0 or 2-19)-SEQ ID NO:21;
- (e) SEQ ID NO:25-[SEQ ID NO:26]_(0 or 2-19)-SEQ ID NO:27;
- 10 (f) SEQ ID NO:31-[SEQ ID NO:32]_(0 or 2-19)-SEQ ID NO:33;
- (g) SEQ ID NO:37-[SEQ ID NO:38]_(0 or 2-19)-SEQ ID NO:39;
- (h) SEQ ID NO:43-[SEQ ID NO:44]_(0 or 2-19)-SEQ ID NO:45;
- (i) SEQ ID NO:49-[SEQ ID NO:50]_(0 or 2-19)-SEQ ID NO:51;
- (j) SEQ ID NO:55-[SEQ ID NO:56]_(0 or 2-19)-SEQ ID NO:57;
- 15 (k) SEQ ID NO:61-[SEQ ID NO:62]_(0 or 2-19)-SEQ ID NO:63;
- (l) SEQ ID NO:67-[SEQ ID NO:68]_(0 or 2-19)-SEQ ID NO:69;
- (m) SEQ ID NO:73-[SEQ ID NO:74]_(0 or 2-19)-SEQ ID NO:75;
- (n) SEQ ID NO:79-[SEQ ID NO:80]_(0 or 2-19)-SEQ ID NO:81;
- (o) SEQ ID NO:85-[SEQ ID NO:86]_(0 or 2-19)-SEQ ID NO:87;
- 20 (p) SEQ ID NO:91-[SEQ ID NO:92]_(0 or 2-19)-SEQ ID NO:93;
- (q) SEQ ID NO:97-[SEQ ID NO:98]_(0 or 2-19)-SEQ ID NO:99;
- (r) SEQ ID NO:103-[SEQ ID NO:104]_(0 or 2-19)-SEQ ID NO:105;
- (s) SEQ ID NO:109-[SEQ ID NO:110]_(0 or 2-19)-SEQ ID NO:111;
- (t) SEQ ID NO:115-[SEQ ID NO:116]_(0 or 2-19)-SEQ ID NO:117;
- 25 (u) SEQ ID NO:121-[SEQ ID NO:122]_(0 or 2-19)-SEQ ID NO:123;
- (v) SEQ ID NO:127-[SEQ ID NO:128]_(0 or 2-19)-SEQ ID NO:129;
- (w) SEQ ID NO:133-[SEQ ID NO:134]_(0 or 2-19)-SEQ ID NO:135;
- (x) SEQ ID NO:139-[SEQ ID NO:140]_(0 or 2-19)-SEQ ID NO:141;
- (y) SEQ ID NO:145-[SEQ ID NO:146]_(0 or 2-19)-SEQ ID NO:147;
- 30 (z) SEQ ID NO:151-[SEQ ID NO:152]_(0 or 2-19)-SEQ ID NO:153;
- (aa) SEQ ID NO:157-[SEQ ID NO:158]_(0 or 2-19)-SEQ ID NO:159;
- (bb) SEQ ID NO:163-[SEQ ID NO:164]_(0 or 2-19)-SEQ ID NO:165;
- (cc) SEQ ID NO:169-[SEQ ID NO:170]_(0 or 2-19)-SEQ ID NO:171;
- (dd) SEQ ID NO:175-[SEQ ID NO:176]_(0 or 2-19)-SEQ ID NO:177;

- (ee) SEQ ID NO:181-[SEQ ID NO:182]_(0 or 2-19)-SEQ ID NO:183;
 (ff) SEQ ID NO:187-[SEQ ID NO:188]_(0 or 2-19)-SEQ ID NO:189;
 (gg) SEQ ID NO:193-[SEQ ID NO:194]_(0 or 2-19)-SEQ ID NO:195;
 (hh) SEQ ID NO:199-[SEQ ID NO:200]_(0 or 2-19)-SEQ ID NO:201;
 5 (ii) SEQ ID NO:205-[SEQ ID NO:206]_(0 or 2-19)-SEQ ID NO:207;
 (jj) SEQ ID NO:211-[SEQ ID NO:212]_(0 or 2-19)-SEQ ID NO:213;
 (kk) SEQ ID NO:217-[SEQ ID NO:218]_(0 or 2-19)-SEQ ID NO:219;
 (ll) SEQ ID NO:223-[SEQ ID NO:224]_(0 or 2-19)-SEQ ID NO:225;
 (mm) SEQ ID NO:229-[SEQ ID NO:230]_(0 or 2-19)-SEQ ID NO:231;
 10 (nn) SEQ ID NO:235-[SEQ ID NO:236]_(0 or 2-19)-SEQ ID NO:237;
 (oo) SEQ ID NO:241-[SEQ ID NO:242]_(0 or 2-19)-SEQ ID NO:243;
 (pp) SEQ ID NO:247-[SEQ ID NO:248]_(0 or 2-19)-SEQ ID NO:249;
 (qq) SEQ ID NO:253-[SEQ ID NO:254]_(0 or 2-19)-SEQ ID NO:255;
 (rr) SEQ ID NO:259-[SEQ ID NO:260]_(0 or 2-19)-SEQ ID NO:261;
 15 (ss) SEQ ID NO:265-[SEQ ID NO:266]_(0 or 2-19)-SEQ ID NO:267;
 (tt) SEQ ID NO:271-[SEQ ID NO:272]_(0 or 2-19)-SEQ ID NO:273;
 (uu) SEQ ID NO:277-[SEQ ID NO:278]_(0 or 2-19)-SEQ ID NO:278;
 (vv) SEQ ID NO:283-[SEQ ID NO:284]_(0 or 2-19)-SEQ ID NO:285;
 (ww) SEQ ID NO:289-[SEQ ID NO:290]_(0 or 2-19)-SEQ ID NO:291;
 20 (xx) SEQ ID NO:295-[SEQ ID NO:296]_(0 or 2-19)-SEQ ID NO:297;
 (yy) SEQ ID NO:301-[SEQ ID NO:302]_(0 or 2-19)-SEQ ID NO:303;
 (zz) SEQ ID NO:307-[SEQ ID NO:308]_(0 or 2-19)-SEQ ID NO:309;
 (aaa) SEQ ID NO:313-[SEQ ID NO:314]_(0 or 2-19)-SEQ ID NO:315;
 (bbb) SEQ ID NO:319-[SEQ ID NO:320]_(0 or 2-19)-SEQ ID NO:321;
 25 (ccc) SEQ ID NO:325-[SEQ ID NO:326]_(0 or 2-19)-SEQ ID NO:327;
 (ddd) SEQ ID NO:331-[SEQ ID NO:332]_(0 or 2-19)-SEQ ID NO:333;
 (eee) SEQ ID NO:337-[SEQ ID NO:338]_(0 or 2-19)-SEQ ID NO:339;
 (fff) SEQ ID NO:343-[SEQ ID NO:344]_(0 or 2-19)-SEQ ID NO:345;
 (ggg) SEQ ID NO:349-[SEQ ID NO:350]_(0 or 2-19)-SEQ ID NO:351;
 30 (hhh) SEQ ID NO:355-[SEQ ID NO:356]_(0 or 2-19)-SEQ ID NO:357;
 (iii) SEQ ID NO:361-[SEQ ID NO:362]_(0 or 2-19)-SEQ ID NO:363;
 (jjj) SEQ ID NO:367-[SEQ ID NO:368]_(0 or 2-19)-SEQ ID NO:369;
 (kkk) SEQ ID NO:373-[SEQ ID NO:374]_(0 or 2-19)-SEQ ID NO:375;
 (lll) SEQ ID NO:379-[SEQ ID NO:380]_(0 or 2-19)-SEQ ID NO:381;

(mmm) SEQ ID NO:385-[SEQ ID NO:386]_(0 or 2-19)-SEQ ID NO:387;
 (nnn) SEQ ID NO:391-[SEQ ID NO:392]_(0 or 2-19)-SEQ ID NO:393;
 (ooo) SEQ ID NO:397-[SEQ ID NO:398]_(0 or 2-19)-SEQ ID NO:399;
 (ppp) SEQ ID NO:403-[SEQ ID NO:404]_(0 or 2-19)-SEQ ID NO:405; and
 (qqq) SEQ ID NO:409-[SEQ ID NO:410]_(0 or 2-19)-SEQ ID NO:411;
 wherein the domain in brackets is an optional internal domain.

The polypeptides of the invention represent novel repeat proteins with precisely specified geometries identified using the methods of the invention, opening up a wide array of new possibilities for biomolecular engineering. The polypeptides of this aspect include 2 or 3 domains, and are represented in Table 1 below, reflected in each row showing listed as “DHRx_variants” (where x is replaced by a specific number in the table). As shown in the table, the residues in brackets are possible variant positions of the residue immediately preceding it. The domains noted as “Ncap” and “Ccap” are always present, while the domain listed as “internal” is optional. When present, the “internal” domain is present in 2-19 copies

TABLE 1.

Module	Ncap	Internal	Ccap
DHR1 variants	G[SDN]C[SDT]D[E]Q[DE]V[I]ET[AK]RE]D[KER]AS[AYR]S[KED]T[RDE]I[V]R[KE]E[NQ]R[V]A[I]AL[E]KQ[K]EN]N[Y]RA]PN[G]Y[ND]S[PA]E[DKT]K[TQD]V[IA]AD[KER]V[EL]AAK[ER]IV[I]K[AL]K[ER]I[V]I[ALV]E[K]G[ERS]N[SRD]P[N]G[G]SDN] (SEQ ID NO: 1)	C[AKN]D[QS]C[A]V[I]AK]A]DR]AAS[ARY]S[A]I[V]R[KE]A]AVI[AL]E[T]K[QE]N[LAF]PN[G]Y[ND]S[PAE]E[DQ]V[A]V[IA]AD[TEI]VAAAIV[I]K[AEL]AI[V]I[ALV]E[KD]G[SQ]N[AS]PN[G]G[SD] (SEQ ID NO: 2)	R]END]D[E]K[C]A[V]I[R]KED]K[AN]AAS[KR]S[A]I[LE]R[KEN]AVQ[KER]E[DKQ]K[QE]N[LAF]P[E]N[G]Y[ND]S[PE]N]E[DKN]V[A]V[KIA]E[KRN]D[I]KT]VK[EHR]R[KDE]AIE[KR]K[DEQ]A[I]R]K[ERQ]E[KDR]G[SAQ]N[AD]PN[G] (SEQ ID NO: 3)
DHR2 design	SDADEAAKEANKAENKARNRNDDEAAKAVKLIKEAIERAKKRNES (SEQ ID NO: 10)	DAVEAAKEAAKALNKAALNRNDDEAAKAVALIAEAIIRALKRNES (SEQ ID NO: 11)	DAVEKAKEAAKALNKAALNRNDDEQAKHVAKQAENIIRALKRNES (SEQ ID NO: 12)
DHR2 variants	S[DET]D[TS]A[S]D[E]E[DKR]AA[KRE]K[RE]E[LAR]AN[D]EQ]K[ER]AE[R]N[K]E[K]LE]A]R[E]N[KRE]R[NKQ]N[G]D[N]D]ES]E]DNS]AA]QIK[K]ER]A]VK[E]L[K]IK[QE]E[RT]AIE[K]T]R]EQ]AK[E]K[ER]R[QK]N[G]E[SD]S[DR] (SEQ ID NO: 7)	D[TE]AV[IL]E[KQ]AAK[AE]E[LRA]AAK[ERQ]ALN[IKQ]K[I]ALN[KQD]R[NQ]N[HGE]D[N]D[ER]E[RD]AAK[ER]AV]A]K[L]KR]IAE]KR]AIIR]EAL]ALK[QER]R[QK]N[G]E[SD]S[DER] (SEQ ID NO: 8)	D]ES]AV[IL]E[KRD]K[RN]A]K[RAQ]E[KQR]AAK[ER]N[K]ET]LN[IKS]K[EQR]ALN]QKD]R[EKN]N[GH]D[SN]D[EQ]E[D]Q[EKA]AK[R]H[KEN]VA]K]K[E]Q[ETR]AE[RK]N[QK]I]R[EKQ]A[D]LK[QR]R[KDE]N[G]E[DQ]S[DET] (SEQ ID NO: 9)

DHR3 design	SSEDTVRKIAQKCSEAIRESN DCEEAARKCAKTISEAIRES NS (SEQ ID NO: 16)	SELAVRIIAQVCSEAIRESND CECAARICAKIISEAIRESNS (SEQ ID NO: 17)	SELAKRIIKQVCSEAKRESN DTECAKRICKIKSEAKRES NS (SEQ ID NO: 18)
DHR3 variants	S[D]S[T]E[D]D[EQ]T[ADE]V[I]R[KQ]K[ERD]I[AV]A[S]Q[K E]K[DQR]C[AV]S[AR]E[KD N]A[D]R[KEQ]E[K]T[S]ENQ[N]K[D]N[C]T[E]DRT]E[KR]A AR[KQE]K[DER]C[A]AK[ET D]T[IEK]I[V]S[RAE]E[KDN]A [LT]I[AT]R[KET]E[KQ]S[AL] N[G]S[N] (SEQ ID NO: 13)	S[TE]E[D]LA[LT]V[I]R[K]I[AV]A[S]Q[AE]V[A]C[AV]S[AR]E[A]AIR[KEQ]E[T]S[A]N D[N]C[T]E[DK]C[AS]AAR[K EH]IC[A]AK[ETR]I[V]S[RAE JE]AKQ[A]L[I]I[AV]R[KE]E[Q R]S[AQ]N[G]S[D] (SEQ ID NO: 14)	S[DEP]E[D]L[K]A[L]R[K]ERD [R]K[Q]I[AV]K[DEN]Q[EA]V[A]C[EA]K[S]REK[E]A[AK]R[R [EKQ]E[TV]S[A]N[K]D[N]T[DEK]E[DK]C[AS]AK[TDN]R[KE]IC[AST]T[KEQ]K[QRE]K[[RE]S[ERK]E[AQR]A[L]K[RE [R]EKN]E[Q]S[NQ]N[G]S[D] (SEQ ID NO: 15)
DHR6 design	SEEKEEALKKVREAAKKLG SSDEEARKCFEEAREWAER TGSS (SEQ ID NO: 22)	AYEAAEALFKVLEAAAYKL SSAEACECFNQAAEWAER TGSG (SEQ ID NO: 23)	AYEAAERLFEELERAYEEGS SAEEACEEFNKKEEAAHRK GKK (SEQ ID NO: 24)
DHR6 variants	S[D]E[D]E[KD]K[DER]E[KN Q]E[TKR]AL[EKR]K[EQN]K[ELQ]VR[E]E[DRT]AAKK[EQ] L[NQ]GS[A]S[N]D[ESQ]E[D] E[QDH]AR[EDK]K[ERQ]C[A NW]F[TW]E[RK]E[RQ]AR[A KS]E[KNQ]W[A]AE[KNS]R[E Q]T[A]GS[AV]S[NDT] (SEQ ID NO: 19)	AY[AW]E[LQR]AAE[HK]AL[A]F[A]K[EQN]VL[A]E[K]AA Y[HAW]K[R]L[N]GS[A]SAE[DR]E[Q]AC[ARL]E[KQ]C[A W]FN[DES]Q[ER]AAE[QKR] WAE[KQS]R[EK]T[N]GS[AV] G[NT] (SEQ ID NO: 20)	AY[AK]E[DO]AAE[HKR]R[E K]L[A]F[A]E[QKR]E[VQ]L[A] ER[EKN]AY[WAH]E[K]EJRN Q[GS]KLE[S]D[AE]RDK[E]Q R]AC[ART]E[KR]E[Q]F[V]N[DS]K[RE]K[ERD]E[AQ]E[KR] E[KR]AH[KQR]R[KE]K[END] GK[QT]K[NDT] (SEQ ID NO: 21)
DHR7 design	STKEDARSTCEKAARKAAE SNDEEVAKQAAKDCLEVA KQAGMP (SEQ ID NO: 28)	TKEAARSFCEAAARAAAES NDEEVAKIAAKACLEVA KQAGMP (SEQ ID NO: 29)	TKEAARSFCEAAKRAAKES NDEEVEKIAKKACKEVA KQAGMP (SEQ ID NO: 30)
DHR7 variants	ST[SD]K[QE]E[DR]D[K]AR[K ET]S[EKR]T[EQ]CE[ERQ]K[RQ]AAR[EQ]K[REH]AAE[K N R]S[QKD]N[KR]D[N]S[E]PK[E [DNK]V[EDQ]AK[ERH]Q[KR E]AAK[REQ]D[ERK]CL[AKR E]R[K]V[A]AK[DOE]Q[KRE] AGM[AL]P[DTN] (SEQ ID NO: 25)	T[RAE]K[R]E[KR]AAR[KEQ] S[EDK]FCE[KQR]AAAR[EK] AAAE[R]S[QEH]N[KR]D[S]E[PKT]E[TKD]V[A]AK[ER]I[V A]AAK[RY]ACL[AKR]E[AQ R]V[A]AK[DEQ]Q[EN]AGM[AL]P[DT] (SEQ ID NO: 26)	T[RPK]K[QR]E[K]AAR[KE]S[ERA]FCE[KR]AAK[E]R[KEQ] AAK[RDE]E[K]S[QKN]N[GK] D[S]E[PDS]E[KQT]V[A]E[KR K]E[R]I[V]A]AK[RED]K[ERQ] ACK[ERQ]E[QAK]V[A]A[KL R]K[ERD]Q[E]AGM[AL]P[DT] (SEQ ID NO: 27)
DHR8 design	SDMEMKVMELKKA VELA KKNNDDEVAREIERAAKEIV EALRENN (SEQ ID NO: 34)	DEMAKVMLALAKAVLLAA KKNNDDEVAREIARAAAEIV ALRENN (SEQ ID NO: 35)	DEMAKKMLELAKRVLDAA KKNNDDETAREIARQAAEEV EADRENN (SEQ ID NO: 36)
DHR8 variants	S[DT]D[STN]E[KDT]M[AIQ] K[EQR]K[EQR]V[A]M[KLR]E [K]A[L]L[W]K[ERD]K[RE]AV [AI]E[QDK]L[Q]AK[SR]K[N QD]N[SD]N[G]D[N]D[EPK]E[DK]V[AQ]AR[KE]E[RKA]IE[KQR]R[KEH]AAK[DEQ]E[R] [A]V[KA]E[E]KDR[AL]A[R]K EN]E[KNQ]N[VA]N[DP]T[S] DQT] (SEQ ID NO: 31)	D[ESK]E[DKL]M[AV]A[WIL] K[ER]V[A]M[AL]L[AEY]A[L] L[W]AK[ELQ]AV[AI]L[AR]L[IQE]AAK[QER]N[SD]N[G]D[N]D[A]E[DK]V[AQ]AR[AIQ] E[RQ]I[AR]KEH]AAA[EK]E[RQ]I[A]V[A]E[RD]K[AL]A[R] AEK[E]KQT[N]VAI[N]TQK[S [DT] (SEQ ID NO: 32)	D[ER]E[DKQ]M[AV]A[WIL]K [DER]K[TED]M[AL]L[RAE]E[KR]L[ERW]AK[EQ]R[KES]V[AI]L[AR]D[RKQ]A[L]AK[QR] N[SDE]N[G]D[N]D[A]E[KD]T [KES]AR[AJK]E[KR]I[QRT]A R[EKD]Q[KEN]AA[EV]E[RK] E[ADK]V[A]E[RKD]A[KNE]D [LAE]R[AKD]E[KRQ]N[G]N[QTE]S[DT] (SEQ ID NO: 33)

DHR9 _design	SYEDEAEKARRVAEKVER LKRSGETSEDEIAEEVAREISE VIRTLKESGSS (SEQ ID NO: 40)	YEVIAEIVARIVAEIVEALKR SGTSEDEIAEIVARVISEVIRT LKESGSS (SEQ ID NO: 41)	YEVKEIVQRIVEEIVEALKR SGTSEDEINEIVRRVKSEVER TLKESGSS (SEQ ID NO: 42)
DHR9 _variants	S[D]Y[STD]E[DT]D[E]E[DT] AE[KR]E[RK]K[RDE]AR[EK] R[KT]V[I]AE[NRD]K[DET]V A[E]K[R]R[KE]LK[YWA]R[KE] D[S]N[KD]GT[V]S[D]E[PNT]D [ET]E[KQ]IAE[KDQ]E[KRT]V [A]AR[EKD]E[QDN]I[VL]S[A RE]E[KR]V[TDK]I[AL]R[KEQ] [T]EDK[LK]EQ[E]KRD[S]RD K]GS[KQ]S[D] (SEQ ID NO: 37)	Y[ESD]E[DKS]V[AED]IAE[K HR]I[V]V[IL]AR[QEA]I[AV]V [I]AE[AKR]I[V]V[A]E[KQR]A LK[QWH]R[EDQ]S[NE]GT[V] S[D]E[PT]D[EQT]E[LQ]IAE[K RD]I[V]V[A]AR[EHI]V[I]I[VL] S[A]EK[E]RV[V]I[IL]R[EKQ] T[AEQ]LK[EQT]E[NKR]S[EQ] N]GS[KQ]S (SEQ ID NO: 38)	Y[SDE]E[DR]V[EQT]IK[RDQ] E[KH]I[V]V[IL]Q[RET]R[EQA] I[AV]V[IAK]E[RKN]E[AKR]I [V]V[EIK]E[KR]ALK[QER]R KE]S[NET]GT[V]S[D]E[PS]D E[E]QLK[I]N[KRE]E[KR]I[V]V [ESA]R[KQ]R[IHO]V[I]K[QR] E]S[EDK]E[KRV]V[AT]E[KR] R[KE]T[AEQ]L[QKN]K[REN] E[KRD]S[QDN]GS[KQE]S[D NP] (SEQ ID NO: 39)
DHR10 _design	SSEKEELRERLVKIVVENAK RKGGDTTEEAREAAAREAFEL VREAAERAGID (SEQ ID NO: 46)	SSEVLELAIRLIKEVVENAQ REGYDISEAARAAAEAFKR VAEAAKRAGIT (SEQ ID NO: 47)	SSETLKRAIEEIRKRVEEAQR EGNDISEAARQAAEEFRKK AEELKRRGD (SEQ ID NO: 48)
DHR10 _variant s	S[T]S[DE]E[DKT]K[AS]E[K]E [KNR]L[IT]R[AKQ]E[KRN]R KE[L]I[V]I[K]I[KT]V[V]AK[E] K[N]AL]AK[QER]R[KE]K[QN R]GD[EQW]D[N]T[EKD]E[SD] E[KD]T[AR]AKE[E]KRD[A]D [AR]KE[E]K]AF[VWA]E[KR] L[R]V[IA]R[EKQ]E[RD]K[AA L]E[KRD]R[EKD]A[S]GI[L]D (SEQ ID NO: 43)	S[T]S[KNT]E[DTK]V[A]L]IA V[E]KQ[L]IT[A]V[I]A[R]KE[L I]I[V]KE]IK[V]V[A]EN]AL]A Q[AW]RE[QKN]GY[EQ]D[N]I [V]S[AT]E[KD]AAR[QEK]A D]AAE[DR]AF[VAW]K[EAQ] R[IQ]V[IA]AE[QR]AA[L]K[E H]R[E]HK]AGI[L]D]T[VDK] (SEQ ID NO: 44)	S[T]S[TKW]E[DKS]T[ADR]L IAV]K[ER]R[EKD]A[V]I]A[E] KD]E[HKD]I[V]R[K]K[EQ]R [E]V[A]E[KQ]E[KRT]AQ[AL] R[KDE]E[KQ]GN[ERQ]D[NT] I[V]S[AT]E[DQK]AAR[EKQ] Q[ERD]AAE[KR]E[KRQ]F[V AW]R[KEA]K[RE]K[EDR]AE QK[E]KRN]L[RA]K[HER]R[K EQ]R[KE]GD[NQK] (SEQ ID NO: 45)
DHR12 _design	DDEEQCREIAEKAKQTYTD DEEIARIIAEAAARQTTTD (SEQ ID NO: 52)	DEEICRCIAEAAKQTYTDDE EIARIIAAARQTTTD (SEQ ID NO: 53)	DEEIERCIEEAAKQTYTDDE EIERIKEYARRQTTTD (SEQ ID NO: 54)
DHR12 _variant s	D[N]D[ST]E[TDQ]E[D]Q[KET]C[A]R[K]E[E]K]IAE[KR]K]QE N]AK[RQ]Q[KR]T[KDR]Y[SA R]T[SD]D[TN]D[PKE]E[DKQ] E[KQA]IAR[KAE]I[ELY]IAE KR[A]E]AR[KHQ]Q[KR]T]EQ R[TTD]N] (SEQ ID NO: 49)	D[PK]E[TD]E[R]IC[A]R[KE]C [L]IAE[R]AAK[RQ]Q[ER]T Y[ASR]T[DES]D[NTS]D[PKE] E[QDT]E[DKN]IAR[AK]I[L]V IAY[AE]AAR[KHQ]Q[KR]T Q]TTD[N] (SEQ ID NO: 50)	D[PES]E[DKN]EIE[RKD]R[K] C[L]IE[K]E[IQ]AAK[R]Q]KE TY[SA]R]T[SD]D[NTS]D[PQ E]E[DN]E[DKN]E[KRD]R[KE Q]I[L]V]K]I[E]KD]Y[IEA]AR EKD]R[KE]Q[EKR]T]QS]TTD [N] (SEQ ID NO: 51)
DHR13 _design	NAEDKAREVLKELKDEGSP EEEAARQVLKDLNREGSN (SEQ ID NO: 58)	AEDAARAVLKALKDEGSPE EEAARAVLKALNREGSN (SEQ ID NO: 59)	EEDASRAVLKALKDEGSPEE EARRAVEKALNREGSN (SEQ ID NO: 60)
DHR13 _variant s	N[SD]A[SDT]E[TAS]D[E]K[K] EDN]AR[ALY]E[K]V[EKQ]L K[EQR]E[TKQ]LK[EQ]D[KR N]E[KQD]GS[TVL]P[SD]E[PT R]E[TRS]E[K]AAR[AEL]Q[K EN]V[L]LK[EQR]D[EKQ]L[N] EK]R[NKE]E[KRQ]GS[V]N[S D] (SEQ ID NO: 55)	A[RTE]E[TIS]D[EQK]AAR[A LY]A[IL]VLK[ERV]ALK[QR N]D[QRK]E[QSR]GS[TVH]P DS]E[PT]E[KST]E[Q]AAR[AL A]ILQ]V[IL]LK[EQR]ALN[EK R]NEQ]E[TNQ]GS[V]N[DS] (SEQ ID NO: 56)	E[TSK]E[DS]T]D[EQQ]AS[AK R]KE]A[IK]VL[EW]K[RQE]A LK[EQR]D[QNE]E[SHQ]GS[V TK]P[SD]E[PR]E[D]E[KR]AR KN]R[EK]A[ILQ]V[A]E[KDR] K[RED]AL]QE]N[KER]R[KN Q]E[TNH]GS[KQH]N[DR] (SEQ ID NO: 57)
DHR14 _design	DSEEVNERVKQLAEKAKEA TDKEEVIEI]VKELAEKQST D (SEQ ID NO: 64)	SELVNEIVKQLAEVAKAETD KELVIYIVKILAEKQST (SEQ ID NO: 65)	SELVNEIVKQLEEVAKAETD KELVEHIEKILEELKKQSTD (SEQ ID NO: 66)
DHR14 _variant s	D[NST]S[DTN]E[D]E[D]V[IE] N[RKE]E[KDN]R[KEN]V[I]K] ERD]Q[KER]L[KR]AE[K]K[E R]AK]Q]E[KR]ATD]NS]K[R T]P]E]DSK]E[KL]V[I]I[KRE]E KR]I[L]V[AL]K]ER]E[KT]L[I]	S[DEN]E[DKN]L[A]V[I]N[KR L]E[KQ]I[A]V[I]K[REQ]Q[LA IL]V]AEVAK]R]E]Q]ATD]NS K]REP]E]DRS]L]V[I]I[REH]Y [ERK]I[L]V[AL]K]RDE]I[AL] L]I[A]ER]E[KQN]LAK]ER]Q]	S[DEP]E[DKR]L[A]V[I]Q]N[K QE]E[RDH]I[A]V]IE]K[EQ]Q LAE]L[V]E[OKR]E[KR]VAFK QR]K[DE]E]Q]ATD]NS]K]DE P]E[DKN]L]V[I]Q]R]E[KR]H]EQ R]I[L]E]NQ]K]ER]I[AL]L]R]

	A[RQ]E[KNR]L[ER]AK[QSE] Q[KR]S[A]T[SNQ]D[NST] (SEQ ID NO: 61)	KDE]S[A]T[QNS]D[NST] (SEQ ID NO: 62)	E[KR]E[KNQ]LK[Q]K[R]Q[R SE]S[ALR]T[NQ]D[KNS] (SEQ ID NO: 63)
DHR15 _design	NDERQKQREEVRKLAELA SKATDEELIKEIKKCAQLAE ELASRSTN (SEQ ID NO: 70)	DELIKQILEVAKLAFELASK ATDEELIKEILKCCQLAFELA SRSTN (SEQ ID NO: 71)	DEEIKQILETAKEAFERASK ATDEEIKEILKKCQEFK KSRSTN (SEQ ID NO: 72)
DHR15 _variant s	N[DS]D[S]E[D]R[ETN]Q[KED] J[K]R[E]Q[L]R[EKQ]E[KQR]E[KIR]V[A]R[E]K[DE]LA[W]E[KR]E[KRD]LAS[KNQ]K[NQR] J[A]L[TEN]D[NS]E[DSP]E[QD] J[L]A[J]R[A]K[DEQ]E[QLR]J[A] J[K]Q[K]E[R]C[A]AQ[KE]L[RK] E]A[W]E[KNQ]E[KDQ]LAS[K NE]R[KQD]S[A]TN[DS] (SEQ ID NO: 67)	D[P]E[TR]L[I]K[IRN]Q[LEA]I [A]LE[K]V[A]AK[IL]LAF[A N]E[K]LAS[QR]K[NER]A[L]T DE[P]E[NR]L[A]J[A]K[E]E[L Q]I[A]L[K]E[R]C[A]C[A]Q[KS] L]E]A[W]F[A]E[K]LASR[K]S[A]TN[D] (SEQ ID NO: 68)	D[P]E[DKN]E[DK]I[K]RAI[Q] REK]I[A]LE[KQR]T[EIK]AK[I L]E[RK]AF[AN]E[KQ]R[KDE] AS[EKQ]K[NRD]A[L]I[T]DE] D[ST]E[DP]S[E]N[KD]E[K]I[A R]K[ES]E[KR]I[A]L[K]E[R]K[E R]C[A]Q[E]E[RKQ]K[REN]F[A]E[KR]K[DER]K[DNS]S[N]R [KQD]S[KN]TN[DS] (SEQ ID NO: 69)
DHR16 _design	NDKAKEAEELLRKALEKAE KENDETAIRCVELLKEALER AKKNNN (SEQ ID NO: 76)	DKAIEAVELLAKALEKALK ENDETAIRCVELLAEALLRA LKNNN (SEQ ID NO: 77)	DKAIEEVLAKALEKALK NDETKIREVCERAEELLRL KNNN (SEQ ID NO: 78)
DHR16 _variant s	N[D]D[T]K[T]A[S]K[DE]E[RD] K[A]E[KQ]E[KD]L[E]K[N]L[R]K DE[K]EDR[AL]E[K]E[RKQ]K[IER]A[E]QR[K]E[R]E[QKR]N[G]D[S]E[DKS]T[KDQ]AI[LQ] R[KE]C[A]VE[K]L[K]L[K]R[OE] J[E]K[Q]ALE[KR]R[E]IL[AK]E[R] K[E]R[N]QRD[N]G[N]D] (SEQ ID NO: 73)	D[E]K[K]ET]AIEAVE[YKR]L[RK]LAK[ED]ALE[RLK]K[IR] ALK[ERN]E[QR]NDE[KS]T[K D]AI[V]R[E]K[C]A]VC[AL]LL AE[R]ALL[E]K[R]EL[ALK]R] N[QER]N[G]N[D] (SEQ ID NO: 74)	D[E]K[DSE]AIE[R]E[TNK]VE [RAL]R[K]E[L]W[AK]ERD[E] KDN]LE[AKL]K[RED]ALK[E RN]E[KNQ]N[G]D[N]E[S]T[D K]K[AQS]I[V]R[E]K[EVC]AL R[E]KR[RAE]K[R]E[KQR]LL[AEK]R[ED]R[AD]L[K]R[E]N[K Q]N[G]N[QK] (SEQ ID NO: 75)
DHR17 _design	SSEDAREKIEQLCREAKEIAE RAKQQNSQEEAREAEKLLR IAKRIAELAKQANQ (SEQ ID NO: 82)	SEVAREAECLCRIAKLIAEL AKQANSQEVAREAEALLRI AKLIAELAKQANQ (SEQ ID NO: 83)	SEVAREAECLSRIAKLIBEL AKQANSQEVKREAQEAldr IQKLIBELQKQANQ (SEQ ID NO: 84)
DHR17 _variant s	S[ND]S[E]D[T]D[EQ]A[N]R[KE] U[E]K[R]K[NDR]I[E]KD]Q[K]E] L[C]LRA]R[KEQ]E[KQR]AK[E Q]E[KR]I[EV]AE[RN]R[EKT] AK[N]QIRKE]Q[SEN]N[GK]S J[N]Q[KR]E]D[E]DQS]AR]IKL J[E]RK]AI[V]E[KRS]K[ERQ]L L[AR]R[KE]I[V]AK[EQR]R[K NQ]IAE[KR]L[E]AK[QRE]Q[KRE]AN[GK]Q[TS] (SEQ ID NO: 79)	S[AP]E[DK]V[A]AR]ALQ]E[R DK]AIE[KR]C[A]L[C]LAE]R[E KH]I[V]AK[RE]LIAELAK]QE R]Q[EN]AN[G]S[D]Q[K]E]DK T]V[A]AR]E]E]RVK]AI[V]E] KDQ]ALL[AR]R[KET]I[V]AK [EQ]LIAE[RK]LAK]Q]Q]DKR JAN[GK]Q[TS] (SEQ ID NO: 80)	S[PAR]E[DKS]V[A]AR]KTE] E]QK]AI[K]E[KR]C[A]LS[KQ N]R[EKT]I[V]A]KE]K[QRE]LI E[KQR]E[R]D]LAK[ERN]Q[E] AN]GK]S[D]Q]DE]E]DKT]V] A]K]RA]R]TKE]E]KIQ]AQ]K E]E]K]AL[AKN]D]EKQ]R]KE Q]I[V]Q]DER]K]Q]L]Q]E]KR E]KQ]LQ]KER]K]R]Q]DEN] AN]GK]Q]ETS] (SEQ ID NO: 81)
DHR18 _design	DIEKLCKKAESEAREARSKA EELRQRHPDSQAARDAQKL ASQAEEAVKLACELAQEHP NA (SEQ ID NO: 88)	DIAKLCIKAASEAAFEASKA AELAQRHPDSQAARDAIKL ASQAEEAVKLACELAQEHP NA (SEQ ID NO: 89)	DIAKKCIKAASEAAFEASKA AEEAQRHPDSQKARDEIKE ASQAEEVKERCERAEHP NA (SEQ ID NO: 90)
DHR18 _variant s	D[STN]I[A]W]E[D]K[D]L[ER] CK[EQR]K[ETH]AE[QKR]S[K EN]E[L]A]AR[DKQ]E[KRQ]A R[KE]S[KED]K[L]RE]AE[QDK R]E[KRS]L[A]R]YKE]Q[KDN] R[QDE]H[RAK]PD[N]G[S]NT] Q]DE]A[V]AR]KNQ]D]LET]A Q]ER]I]K]E]L]A]V]A[V]S]EKR] Q[AEL]AE[KQ]E[RKQ]AVK] ER]L[E]KQ]ACE[KNR]LAQ[K N]E[KOR]H[Y]P[K]N[G]A[S] (SEQ ID NO: 85)	D[EQ]I[A]AK[LQR]L[RK]CI] L]K]ET]AAS]AIQ]E[LAR]AA E[KRI]AAS[AK]K[L]AQ]AA[I J]E[KDS]L[A]A]L]Q]KLR]R[D QE]H[RAL]PD[N]S]NT]Q]ED K]A[V]AR]KAE]D]LEK]AI[L] K]ERQ]L]A]V]A[V]S]AIR]Q]A LE]AAE[KQR]AVK]YLQ]L[E KQ]ACE[KRQ]LAQ]E]E[KQR JH]Y]PN]G]A[S] (SEQ ID NO: 86)	D[EQ]I[A]EQ]AK[R]K]RED] CI]L]K]E[R]A]DKE]AS]AE]E] KR]AAE[KR]E[ANQ]AS[A]E] K[R]E]AA]I]E[QDR]E]ILK]A[L]Q[KRS]R[KDE]H]RY]PD[N] S]D]T]Q]EDS]K]DER]AR]KE Q]D]KER]E]AKD]I]L]K]EDR] E[KRQ]A[V]S]RAI]Q]EKR]K[DLT]AE[RD]K]E[KRD]VK[LA I]E[RKQ]R[KDE]CE[KR]R]K] AQ]ED]E]KQ]H]NY]PN]G]A] SQ] (SEQ ID NO: 87)

DHR19 _design	DEIEKVREEAEKLLKKKTTDE DVLEVAREAIRAAKEATS (SEQ ID NO: 94)	DEILKVIKEALKLAKKTTDK DVLEVAREAIRAAEEATD (SEQ ID NO: 95)	EEILKEIKEALKKAKETTTDT EELEKAREQIRKAEESTD (SEQ ID NO: 96)
DHR19 _variant s	D[TS]E[DKN]I[KQ]E[KQD]K[EHQ]V[A]R[IK]E[KDN]E[DR] AE[KQN]K[ER]L[V]K[SRA] K[RDE]K[QT]TD[NT]D[T]E[Q] D[D]EN[V]A[L]QKR[E]RKD] VAR[KDE]E[LAK]AI[K]R[EK] JAAK[ED]E[NDK]ATS (SEQ ID NO: 91)	D[SEQ]E[DKN]ILK[ERT]V[A] IK[EQ]E[Q]ALK[R]L[IV]AK [QSE]K[QST]TTD[T]K[TED]D [EN]V[A]LE[KR]VAR[ELQ]E[QKL]AIR[EK]AAE[RT]E[ND K]ATD[S] (SEQ ID NO: 92)	E[DSK]E[DS]ILK[EQ]E[RKL]I K[QEN]E[RKN]ALKK[IRE]A K[QS]E[TKQ]TTD[T]TEKS[E [D]E[VD]LE[KRN]K[ER]AR[E KL]E[K]Q[TED]IR[EQ]K[D QR]AE[RT]E[KNQ]S[EQ]TD [N] (SEQ ID NO: 93)
DHR20 _design	SDIEEIRQLAEELRKKSDNEE VRKLAQEAELAKRSTD (SEQ ID NO: 100)	SDVLEIVKDALELAKQSTNE EVIKLALKA AVLAAKSTD (SEQ ID NO: 101)	EEVLEEVKEALRRAKESTD EEIKEELRKAVEEAESTD (SEQ ID NO: 102)
DHR20 _variant s	S[TDN]D[TQ]I[VAR]E[KD]E[KR]IR[EIQ]Q[EK]L[TEK]AE [RKQ]E[RQD]L[V]R[ASK]K[NRT]K[EDN]S[ALT]D[T]N[D] E[DPK]E[TDQ]V[A]R[IQ]K[RED]LAQ[ERK]E[RTL]AAE[K]LAK[HQ]R[K]S[ANT]T[R] D[TS] (SEQ ID NO: 97)	S[KEP]D[TKQ]V[A]L[W]E[K R]IVK[EQ]D[LKR]ALE[KR] L[VT]AK[EQ]Q[KRD]S[AT]T N[D]E[DPN]E[DK]V[A]IK[R A]LALK[ELR]AAVLAQK[QR S]AEN[T]R[D]TS] (SEQ ID NO: 98)	E[KPS]E[DKT]V[A]L[W]E[K N]E[TIR]VK[ERA]E[KR]ALR[EQ]R[KDE]AK[EQ]E[KR]S[AKN]TD[N]E[DN]E[DQR]E[KDN]IK[RAE]E[ERKQ]E[ADL] LR[EK]K[NQR]AVE[RD]E[D QA]AE[KQ]S[KRT]T[NR]D[IT N] (SEQ ID NO: 99)
DHR21 _design	SEKEKVEELAQRIREQLPDT ELAREAQELADEARKSDD (SEQ ID NO: 106)	SEALKVVYLALRIVQQLPDT ELAREAELAKEAVKSTD (SEQ ID NO: 107)	QEALKSVYEALQRVQDKPN TEEARESLERAKEDVKSTD (SEQ ID NO: 108)
DHR21 _variant s	S[DTN]E[KDL]K[AQS]E[K]K[EDR]VE[R]E[KQS]LAQ[REK] R[KDE]I[V]R[AK]E[KN]Q[NT L]P[K]D[N]TE[DRS]L[I]AR[E K]E[LKQ]AQ[ENL]E[KRQ]L[V]AD[EK]E[KDQ]AR[KEQ] K[ERT]SD[NTR]D[SN] (SEQ ID NO: 103)	S[EQD]E[KNQ]AL[W]K[E]VV [A]Y[KAE]LALR[QAE]I[V]V[A]Q[EK]L[Q]RT]LPD[N]TE[D Q]L[I]AR[KE]E[KLD]ALE[KR D]L[V]AK[EQ]E[KDN]AV[I] K[ER]ST[Q]D[SN] (SEQ ID NO: 104)	Q[EDK]E[DKR]AL[W]K[ED]S [IKD]V[A]Y[KAE]E[KQR]AL Q[EK]R[RTD]V[A]Q[EK]L[D] KQR[K]YHR[PNT]E[D]E[DK] AR[KEQ]E[KQR]S[A]LE[DQR R]KEQ[AK]EQ[E]K[D]EKA [V]IA[K]ET]S[R]T[NQ]D[NST] (SEQ ID NO: 105)
DHR22 _design	DDAEELRERARDLLRKNGS SEEEIKKVDEELEKIVRKAD S (SEQ ID NO: 112)	DDAVKLAVKAAALLAENG SAEEIVKVLEELLKIVEKAD S (SEQ ID NO: 113)	EEEVKDAVREAAELAERGS SAEEIRKQLKDRLRKVEESD S (SEQ ID NO: 114)
DHR22 _variant s	D[S]D[TK]AE[D]E[KT]LR[A] E[QK]R[KL]AR[A]D[KQE]LL R[KQ]K[DEQ]NGS[AQ]S[D]E [DKP]E[DS]E[QS]IK[N]K[RQ] VD[LT]E[K]E[ADL]L[I]E[KQ R]K[RQ]I[A]V[RK]R[DEK]K[QDN]AD[QK]S (SEQ ID NO: 109)	D[SW]D[KET]AV[A]K[ITA]L AV[A]K[L]AAALLAE[QKR]N GS[AQ]SAE[QDS]E[Q]IV[RA Y]K[R]VLE[H]E[ALW]L[I]L A[K]R[I]A[V]I[E]QK]K[Q]AD [Q]S (SEQ ID NO: 110)	E[SW]E[DKS]E[QT]V[A]K[IT A]D[REK]AV[A]R[KEL]E[TD K]AAE[DQ]L[QER]AE[QKR] R[KDE]GS[RE]SAE[DRS]E[R] IR[AY]K[E]Q[TES]LK[EHR]D [EKN]R[L]Q[L]AE[R]KEQ[K] D[V]ILT]E[QKR]E[KNQ]S[A] D[QT]S[D] (SEQ ID NO: 111)
DHR23 _design	SDSEKLAKRVLKELKRRGTS DEELERMKRELEKIKSATS(SEQ ID NO: 118)	SDAMRLALRVVLELVRRGT SSEILEKMMRMLIKIISATS (SEQ ID NO: 119)	DDQMREALRVLEEVRRGT SSEQLERSMRKLIKIKRTS (SEQ ID NO: 120)
DHR23 _variant s	S[TDN]D[TR]S[AQ]E[DK]K[E QR]LAK[QRD]R[EXT]V[A]L [VR]K[ENR]E[QDL]L[A]K[R] R[KN]R[NKS]CT[QE]S[D]D[S P]E[DT]E[DAI]L[E]E[KNR]R [K]M[ALI]K[ER]R[EOK]E[LA Q]L[I]E[KQR]K[RDE]I[VL]I[R KQ]K[DER]S[EQ]T[AT]Q[S]T (SEQ ID NO: 115)	S[DE]D[TEK]AM[A]R[KEA]L ALR[EK]V[A]V[L]I[LE]RQ]L[A]V[A]R[KE]R[KN]GT[EKQ] SS[AIQ]E[DR]T]I[EAN]L[I]E[DKS]K[RT]M[ALI]M[A]R[EK M]LAQ[L]I[I]QR[K]ERO[I]V L]IQ[EK]S[EQ]A[AT]QK]S[T] (SEQ ID NO: 116)	D[ES]D[ET]Q[EAL]M[A]R[K AE]E[RKQ]ALR[KE]Q[ETD]V [L]I[LE]DRK]E[ADR]V[A]IR[KEQ]K[ETD]GT[KQR]S[D]S[AIQ]E[DQR]Q[EDS]L[I]E[KD R]R[KEQ]S[TE]M[A]R[EQ]K [EQ]L[I]I[KQ]K[RE]E[K]K[Q R]K[NDQ]R[S]T[Q]S[DT] (SEQ ID NO: 117)
DHR24 _design	SEAEELARRAAKEAKELCK RSTDEELCKELKKLAELLKE LAERYPD (SEQ ID NO: 124)	SEAAKLALKAALAEALCKQ STDEELCEELVKLAQKLIEL AKRYPD (SEQ ID NO: 125)	SEEAKRALKEAKELIEOCKE STDEDECRELVKRAEELIRE AKENPD (SEQ ID NO: 126)

DHR24 _variant s	SE[QDQ]AE[KQ]E[KOR]L[E] AR[E]R[EK]AA[EK]K[E]E[RK] A]AK[REQ]E[KQS]L[AV]CK R[DK]E[S[KTQ]T[NR]D[N]E[D] E[DKR]L[T]CK[E]E[DKL]LK [EQ]K[ER]LAE[KQR]L[EQQ] LK[EN]E[KQR]LAE[KRD]R[KEN]Y[L]PD (SEQ ID NO: 121)	SE[RTD]AAK[ERQ]LALK[RE] S]AAL[AK]E[AKR]A[L]E[KR] H[L]AV]CK[REQ]Q[EKD]S[Q T]T[N]D[N]E[DNS]E[DKN]LC E[RQ]E[KL]LV[A]K[ER]LAQ[KES]K[ELQ]L[V]A[E]KR]LA K[EQD]R[EK]Y[L]P[S]D (SEQ ID NO: 122)	SE[D]E[ANQ]AK[ERQ]R[EK] ALK[ER]E[NRK]AK[AEL]E[K RN]L[A]I[L]E[RK]Q[EK]R[CK] RQE]E[KQR]S[DK]T[D]D[N]E [DTS]D[EQQ]E[KR]CR[EQQ] E[KR]LV[A]K[ER]R[KEQ]AE[KQ]E[KR]L[EDK]I[V]A[R]KE] E[KR]AK[EQ]R[E]KD[N]DH]P D[K] (SEQ ID NO: 123)
DHR25 _design	DERDKVRELIDRVEKELKRE GTSEELIEEIRKVLKKAKEA ADSDD (SEQ ID NO: 130)	DEAIKVAKEIVRVILELVRE GTSSLEIEILKVLSLAAEAA KSTD (SEQ ID NO: 131)	EEAIKKAKEIVRRILELTREG TSEEEIREELKELRKKAAQKA KSPE (SEQ ID NO: 132)
DHR25 _variant s	D[T]E[DK]R[A]D[KE]K[E]V[A]R[KS]E[K]LID[EQQ]R[EK Q]V[A]E[KR]K[E]E[QL]LK[Q E]R[K]E[RSQ]GT[EQK]S[D]E [SPD]E[DN]LIE[KTN]E[QA D]IR[Q]K[ER]VLK[DRT]K[LE N]AK[QDE]E[KQS]AAD[NKR]SD[S]D[N] (SEQ ID NO: 127)	D[E]E[KD]AIK[E]V[A]AK[QY E]E[L]V[A]R[EKD]V[A]IL[A KR]E[L]R[L]V[AT]R[EK]E[SO R]GT[EQQ]S[D]S[P]E[KRS]LI E[QKR]E[QKD]ILK[ER]VLS[AEK]L[EK]AAE[KLR]AAK[N RA]S[A]T[SP]D[N] (SEQ ID NO: 128)	E[DR]E[DS]AIK[RE]K[IEQ]A K[R]YE]E[KR]IV[A]R[EKD]R[T]IL[AKR]E[R]LT[VAS]R[QK E]E[RKD]GT[EQ]R[S]DNT]E[S P]E[DN]E[KDQ]IR[SEK]E[K] E[TQR]LK[E]E[KQ]LR[AEK] K[E]K[REQ]AQ[KER]K[E]AK [ANR]S[K]P[S]E[D] (SEQ ID NO: 129)
DHR26 _design	DECERLRQEVEKAKELEK LAKQSTDEEVRQIAREVAK QLRRLAEEACRSNS (SEQ ID NO: 136)	DECLRLASEVVKAVQELVK LAEQATDEEVIRVALEVARE LIRLAQEACRSND (SEQ ID NO: 137)	EECLREASEVVKVQELVK EAEKSTDEEEIRELLQRAEE RIREAQERCREGD (SEQ ID NO: 138)
DHR26 _variant s	D[NT]E[DK]CE[KD]R[KE]LR[NQ]Q[EK]T[E]ADK]VE[KDQ] K[RS]AE[QKI]K[JED]R[E]ALK] LE[NKQ]K[ERD]L[V]A[A]K[K] [RDQ]Q[KNE]S[A]T[N]D[N]E [P]E[ND]R[V]AIL[R]I]Q[KNR] I[LEK]AR[KQ]E[KTD]VAK]E D]Q[EAL]LR[EQQ]R[EQQ]LA E]RDK]E[LDH]ACR[KN]S[N QE]N[G]S[D] (SEQ ID NO: 133)	D[K]P[E]E[DNK]CL[I]R[KE]LA S[EK]R[E]QR]VV[A]K[EQ]R[A V]A]Q[KER]E[LKA]LV[A]K] EDQ]L[V]A]AE[KRA]Q[KNE] A[S]TDE[P]E[KNQ]V[A]IL]IR[KE]V[L]IK]AL[A]E[RDK]VAR [AEL]E[LAR]LIR[EKN]LAQ] YAL]E[L]IK]ACR[EK]S[QNE] N]GR]D[N] (SEQ ID NO: 134)	E[DKP]E[E]NSD]CL[I]R[KEN]E [T]AS[EAY]E[KQ]VV[A]K[E QR]E[RKS]V[A]Q[KER]E[K]L V[A]K[EQ]E[KQR]AE[KLR]K [R]S[AD]TD[N]E[P]E[NDQ]E] KRS]IR[K]E[KR]L[AD]L[A]Q] KER]R[EQQ]AE[ALQ]E[KRD]R[EQ]T]IR[KEN]E[K]AQ[EA Y]E[K]R[KNQ]CR[KEQ]E[KN R]GD]Q] (SEQ ID NO: 135)
DHR27 _design	TRQKEQLDEVLEEIQRLAEE ARKLMTDEEEAKKIQEEAE RAKEMLRRAVEKVTD (SEQ ID NO: 142)	NEVIEKLLEVVKELIRLAEEA MKKMTDEEEAAKIAKEALE AIKMLARAVEEVTD (SEQ ID NO: 143)	KERIEQLLREVKEEIRRAEEE SRKETDDEEAAKRAREALR RIRERAREVEEDKS (SEQ ID NO: 144)
DHR27 _variant s	T[SD]R[EDK]Q[ATV]K[ED]E[KR]Q[REK]L[IA]D[KR]E[QT] V[A]L[IVE]E[K]E[R]I]Q[KR]R[KE]L[A]AE[DK]EAR[A]K[RQ]L[RK]M[AE]T[SD]D[SNT]E] DPS]E[NDK]E[KQR]AK[NQ] K[ER]I]Q[K]E[KDN]E[QDK]A E[K]R[KEQ]AK[IQ]E[KQR]M [ADL]L[IT]R[KED]R[DQE]A V[SAH]E[KR]K[QE]V[I]T[DE] D[N] (SEQ ID NO: 139)	N[VAD]E[EDQN]V[AL]I[L]V[E] KQR]K[ERQ]L[IA]L[A]I]E[KH R]V[A]V[A]K[ERQ]E[RL]IR [E]L[A]AE[QK]E[RK]AM[A]K [ER]K[L]R[M]A]T[ES]D[NT]E] KDP]E[QK]E[EDQ]R]AAK[ER]I A]I]K[ARE]E[KQ]ALE[KQR] AIK[A]M[ADL]L[IQ]AR[AE] AV[A]E]IK]E[QD]V[I]T[Q]D] N] (SEQ ID NO: 140)	K[NDE]E[EDN]R[KQD]I[L]V[E] KR]Q[KRE]L[IAT]L[A]I]R[ED K]E[KQR]V[I]A]K[ELN]E[KR] E[I]IR]KE]R[EKN]AE[KQR]E] QRK]E[RKD]S[A]R[KED]K]R E]E[A]T[DS]D[NST]D[KPR]E] QN]E[KDR]AAK[ERN]R[I]E]A [I]R[KAL]E[KQR]ALR[QEK] R[KDQ]IR[AK]E[KQN]R[ETH]AR[KND]E[KRD]V[AE]E[RQ K]E[KR]D[EKR]K[TDQ]S[DN G] (SEQ ID NO: 141)
DHR28 _design	DEEVQRIRREEVRRRAIEEVRE SLERNDSSEAEELAREALER VAEEVKESIKERPDR (SEQ ID NO: 148)	DLAIEAIRALVRLAIEIVRLA LEQNDSLEAREVAEEALRA VAEEVKELAIRQRGDR (SEQ ID NO: 149)	ELAKEAIRALRRLAEEIRRL AAEQNDDELAREVEELARE AIEEVKELERQRPGR (SEQ ID NO: 150)
DHR28 _variant s	D[TN]E[D]E[DNQ]V[IRK]Q[E] KR]R[KN]I[AL]R[KE]E[NKQ] E]TQ]V[A]R[KE]R[KQE]A]A VK]E[RKQ]E[DKQ]V[A]R]K	D[EQ]L[IVE]A[EKQ]E[KQ]A I[AL]R[KE]A[V]LV[A]R[EK]L [AT]A]V[AE]E[RQ]I[AL]V[IA R]KEQ]L[E]ALE[KDQ]QN]G	E[DS]L[IVE]AK[ED]E[KRD]A I[LEA]R[KQ]A[L]V]LR[EKI]R] E]L[AT]AE[RK]E[RT]I[AL]R] IVA]R[KN]L[E]AE[KQE]KQ

	EQ[E]KDR]S[A]LE[DKR]R[E KN]N[G]D[N]S[PT]E[D]E[K]A E[ALK]E[K]L[IR]AR[EKQ]E KNQ]ALE[KDR]R[KTQ]V[I]A E[RQK]E[Q]A[V]A[K]R]E[RK]S[ATI]IK[RQ]E[KNQ]R[HAK]PD[NG]R[TS] (SEQ ID NO: 145)	ID[N]S[P]E[DKQ]L[V]AR[EL A]E[RKN]V[IA]AE[KQR]E[K T]ALR[KE]AV[I]AE[QS]V[A] V[A]K]Q]EA[I]IR[K]QR[A]G P[D]N]R[T] (SEQ ID NO: 146)	D]Q[H]N[G]D[N]D[PSQ]E[DK R]L[V]AR[EKQ]E[RKN]V[IA] E[KR]E[RK]L[EQN]AR[EDK] E[RKQ]A[V]E[KNR]E[R]V[A]R]QED]K[ERN]E[QTV]L[RE K]E[KR]K[EN]Q]E[R]A]PG[N]R[T] (SEQ ID NO: 147)
DHR29 _design	SEVEESAQEVKRAQEVREE AERRGTSQEVLEIKRVVDE ARQLAQRAKESDD (SEQ ID NO: 154)	SEVAESALQVVREALKVVL SALERGTSEEVLKEILRVVS EAIKLALAEIKSSD (SEQ ID NO: 155)	SETARRALEKVRESLKEVLE QLERGTSEELRESLREVSE NIRKALEEIKSPD (SEQ ID NO: 156)
DHR29 _variant s	S[TD]E[DKR]V[ALT]E[KR]E KQT]S[ALQ]AQ[RED]E[KRQ]V[A]E[KQ]K[DE]R]ELA]A L]Q[KED]E[KRN]V[AL]R[EI K]E[KQ]E[QRD]AE[KRQ]R[K DE]R]QTE]GT[V]SQ[SDP]E[D]V[AT]L]IQV]D[KRN]E[QDA] IK]QEIR]KEQ]V[A]V[IA]D[K ER]E[DKL]A[VL]R[KQE]Q[E R]L[AV]A[ILV]Q[KE]R]ELQ] AK]RNE]E[K]S[AL]D[STQ]D] NS] (SEQ ID NO: 151)	S[QEK]E[DKR]V[A]AE[K]A]S AEK]ALQ]EKR]V[A]V[IAL]R [AEK]E[ALK]A[L]L[W]K[QL R]V[AL]V[ALI]L[QR]S[QEA] ALE[KQR]R[QET]GT[V]SE[D RW]E[DK]V[A]L[IV]K[RAD] E[LKD]L[I]R[EK]T]V[A]V[IA] S[KQA]E[RLN]A[V]I[L]K[RE] L[AV]A[ILV]L[EKQ]E[QIK]A I[L]K[NRD]S[AL]S[T]D[NS] (SEQ ID NO: 152)	S[QER]E[DK]T[EDQ]AR[EKL]R[KED]ALE[KR]K[ED]V[IA L]R]AKE]E[KR]S[AL]L[W]K] ENQ]E[KDQ]V[ALI]L[Q]A]E KQR]Q[ADR]L[Q]E[KRN]R[K DQ]GT[KER]SE[DPR]E[DK]E [QDK]L[IV]R[AKN]E[K]S[I]Q T]L[I]R[KE]E[KR]V[IA]S[EK Q]E[KR]N[RTV]I[L]R[KEN]K [QRE]A[ILV]L[EIK]E[KR]E[D NK]I[L]K]RNQ]S[R]P[ST]D[S] (SEQ ID NO: 153)
DHR30 _design	STVKELLDRAELMRELAE RASEQGSDEFEARKLLEDLE QLVQEIIRRELEETGTS (SEQ ID NO: 160)	SEVIRLIAKAIMLMAELALR AAEQGSDAEEAMKLLKDLL RLVLEILRELRETGTD (SEQ ID NO: 161)	KEEIRKVAEEIMMRAKTALD EARQGSDAEEAMKRLKEQL RRILERLREEREKGTD (SEQ ID NO: 162)
DHR30 _variant s	S[NT]T[DEK]V[AIT]K[ED]E KRN]L[A]L]E[D]KNE]R]KE] AR[KEL]E[K]L[R]M]AL]R]E K]E[KQ]L[A]AE[KRD]R]QEL]AS[AKR]E[KR]Q[ED]GS[AQ N]D[TN]E[PSK]E[DKN]E[RK] AR[KNK]K]EQ]LLE[KD]D]E K]L[IV]E[QKR]Q[ERK]L[A]V]IE]Q[KED]E[RDK]R]QKN]R [EK]E[ALQ]L]E[DK]E[DKR]T] SA]GT[A]S[TD] (SEQ ID NO: 157)	S[NKD]E[KNR]V[A]I[KAQ]R] KIE]L[A]I]V]AK]E[AIM]A[L M]AL]AE[KQR]L[A]AL]AV]R [EKL]AAE[KDR]Q[ED]GS[A Q]D[NT]AE[AK]E[KR]AM[A L]K]Q[LLK]R]I]D[EK]L[IV]LR [E]L[A]V]I]L[A]E[RK]IL[I]R] EQD]E[AL]LR]KET]E[KR]T] AS]GTD]ST] (SEQ ID NO: 158)	K[DPQ]E[DPQ]T]E[QT]I[KAQ] R[KE]K[REN]V]TDN]AE[KR]E[KRT]M]A]R]DE]R]ALK]A K]REQ]T]EQD]AL]VKN]D]E RK]E[RQK]AR[KED]Q[KDR] GS]EQD]D[NT]AE[KAR]E[K DQ]AM]AL]K]EQ]R]EKN]LK [RL]E[K]Q]E[KR]L]R]EK]R] KN]I]Q]T]L[A]E[RK]R]EK]L[I] R]EKQ]E[KNR]E[RKL]R[K]E] KDN]K]QEN]GTD]T] (SEQ ID NO: 159)
DHR31 _design	DSYTERARKAVKRYVKEEG GSEEEAEREAEKVREEIRKK ASD (SEQ ID NO: 166)	SYLIQAAAAVVAYVIEEGGS PEEAVKIAEEVVRRIKEKAD D (SEQ ID NO: 167)	RELIRRAAERVAEVIERGGS PEEAVKEAEKVKKQKEES D (SEQ ID NO: 168)
DHR31 _variant s	DS[D]Y[A]T]E]E[KR]R]QEK] AR[AN]K]ERD]A[L]V[A]K]A] E]R]KDE]Y]W]V]AT]K]ERQ] E[K]E[KQ]GG]QY]S]T]E]P]E] D]E]QR]AE[KR]R]KE]E[DIN] AE[KN]K]ER]V[L]R]VEK]E] KQR]E[KR]I]AL]R]EK]K]DN Q]K]QE]AS]NDE]D (SEQ ID NO: 163)	S[DQ]Y[A]L]L]Q]ER]AAAA V[A]V[A]I]AY]W]V[A]I]L]E]K N]E[KQ]GG]QY]S]D]T]PE]D]E [DR]AV]A]K]RE]I]ERQ]AE[R S]E[KQR]V[L]VR]EK]R]KE]I] AL]K]EQ]E]KNT]K]NQ]AD[N R]D(SEQ ID NO: 164)	R[D]E]DS]L]L]R]EKQ]R]KQ S]AAE[KQR]R]QEK]V[A]I]AE [RDK]V]AEQ]I]L]E]KR]R]KQ N]GG]KNQ]S]T]PE]D]E]QKR] AV]A]K]RO]E]N]AE]K]K]RE] E[LQR]VK]R]K]ER]Q]ED]K]E Q]E[KNQ]E[KDN]S[R]D]TN] (SEQ ID NO: 165)
DHR32 _design	SIQEKAKQSVIRKVKKEEGGS EEEARERAKEVEERLKKEA DD(SEQ ID NO: 172)	STLVRAAAAVVLYVLEKGG STEEAVQRRAREVIERLKKEA SD (SEQ ID NO: 173)	EELIREAAKEVLKVLEEGGS VEEAVARERERIEELQKRSD D (SEQ ID NO: 174)
DHR32 _variant s	S[DT]I[TAQ]Q]E]E[DKQ]K[R DQ]AK[A]Q[NRE]S[A]V]R]R [KE]K[WY]V]AE]K]QER]E[K QN]E[QKR]GG[KY]S]ND]E]D]E]D]E]KQ]AR[KQ]E[KRN]R]	S[D]T]AQ]LV]I]KA]R]KL]AA AAVVL]YAW]Y]WA]V[A]LE [QK]K]EQ]GG]Y]S]ND]T]V]E [D]E]T]AV]IL]Q]KRE]R]IKQ] AR]EK]E]QR]V]AT]E]K]R]R]	E[D]E]DSK]L]EQ]I]VKA]R]KI N]E[KST]A]AK[NQR]E[VOR] VL]YAW]K]EQ]V]AT]LE]DN Q]E[RKD]GG]Y]S]ND]V]T]E] DQ]E]Q]AV]IL]E]KTD]R]E]A]

	EKLJAKJEJERKQJ[AT]EJR QJERKQJ[DE]LJIKRQJ RTDJEKNSADJKNEDJSTJ (SEQ ID NO: 169)	DKNLJIKJEQJKNTEJQODJ ASJNDKJDJSJ (SEQ ID NO: 170)	RJEJJEKQJ[QEA]E[RKJE [KRJLERDQJEKSJKTENJR KENJ[ARJDJTNJD (SEQ ID NO: 171)
DHR33 design	SETEEVKKLVEEKVKKEGG SPEEAKETAKEVTEELKEES QD (SEQ ID NO: 178)	STLLKVAALVASAVLKEGG SPEEAAETAKEVVKELRKSA SD (SEQ ID NO: 179)	EELLKEAARQAEESLRQGS PEEAAEEAKKEVKKLKEKS QD (SEQ ID NO: 180)
DHR33 variant s	S[DTN]E[AL]TJ[ELS]E[K]E[K D]VK[AN]K[ER]L[R]V[A]E[A K]E[KRQ]K[WQA]V[AT]K[Q R]K[NDQ]E[QRD]GG[KQ]S[N D]P[D]E[D]E[RQ]AK[QE]E[K QR]T[EKL]AK[DR]E[RK]V[A J]T[VER]E[KD]E[RKD]LK[RT] E[RKT]E[AQN]S[A]Q[DHR]D JSTJ (SEQ ID NO: 175)	S[DE]T[LAE]LL[AR]K[EQR] VAALV[A]AS[AK]A[WEL]V A]LK[DE]E[QDK]GG[Q]S[NT D]PE[D]E[Q]AA[V]E[KR]T[K QE]AK[ERA]E[R]V[A]VK[RD E]E[RKD]LR[TK]K[DER]S[Q AT]AS[QH]D (SEQ ID NO: 176)	E[D]E[KDN]LL[AR]K[EQ]E[K RD]AAR[KEQ]Q[VRE]AE[A] E[KRQ]S[VAT]LR[KEQ]Q[RK D]GK[GG]S[NTD]PE[D]E[QR] AA[V]E[KR]E[NR]AK[AER]K [RE]E[QHR]VK[ER]K[EQR]L ENQJ[TNQ]E[KNR]K[RE]S Q[T]D[K] (SEQ ID NO: 177)
DHR35 design	SEDEVAQASRYAKEQGG DPEKSREEAEKALEEVKKQ ATS (SEQ ID NO: 184)	SEALQVALEAARYASEEGE DPAEALKEAARALEEVRRS ATS (SEQ ID NO: 185)	EEDLKEALDRAREASERGQ NPAESLKEAAEELKKKKEK SSD (SEQ ID NO: 186)
DHR35 variant s	S[NT]E[DT]E[QKR]D[E]KQJ KQJ[V]A[A]K[RE]QJ[ELW]AS [A]R[E]KDY[W]AK[SQR]E[K NR]QGG[QH]D[N]PE[N]K[ED Q]S[A]R[LK]E[K]E[KDR]AE KRN[K]ER[ALE]K[E]LQJ[V[A J]K[REN]K[R]Q[A]AT[QS]S[T] (SEQ ID NO: 181)	S[D]E[D]A[D]L[E]KJQ[KR]V A[ALE]LW[AAR]KE[Y]W[AS [YHR]E[KNQ]E[Q]GE[Q]D[N] PAE[DK]ALK[EQR]E[R]AAR KE[ALE]K[E]QK[V]A[R]KN R[K]S[A]AT[E]S[T] (SEQ ID NO: 182)	E[D]E[D]D[A]L[E]KJQ[E]E KRQ[ALD]KERJ[E]AR[KDE J[E]R]AS[YAQ]E[KNQ]R[ED Q]GQ[E]N[D]PAE[DQ]S[A]L K[E]HQ[E]KRQ[A]E[KR]E[K R]LK[E]K[E]Q[R]K[ER]Q[K]SN E[KR]K[E]SS[TQ]D[RT] (SEQ ID NO: 183)
DHR36 design	SDLEKALKRFVKEEKKGR NPPEAKKEAKKLKKLKKKS AGS (SEQ ID NO: 190)	SDLLTALAKFVLEEVVKGR NPPEAVKEAIKLAELKRSA GS (SEQ ID NO: 191)	SEQLKELATKVLEEVKKGR NPKRAVEEAQKAKEDRKR SNS (SEQ ID NO: 192)
DHR36 variant s	S[T]D[EN]LE[DK]KALK[NEQ J]RQEN[FY]VJIK[RED]E[D QK]E[Q]K[ET]KJ[DR]E[KGR]Q KJ[N]DTSJ[P]ER[E]DKQJ[E]KD QJAK[R]K[RED]E[SD]AK[ER J]K[E]LK[ER]K[ER]K[RD]LK RE[K]RNT[S]A[AGS] (SEQ ID NO: 187)	S[D]D[AKN]LLT[KEQ]ALAK [TDR]F[Y]VLE[QD]E[Q]VRK EQJ[KGR]KQJ[N]TD]PEE[K]A VK[R]E[S]AIK[E]LAE[Q]K[N R]LK[RQ]R[KN]S[A]AGS (SEQ ID NO: 188)	S[D]E[DQ]S[Q]E[LE]RKT[K]E LAT[KRE]K[ESH]VLE[K]E[R AL]VK[Q]E[K]R[GR]EQT[N]T D]PK[E]R[EDK]AVE[RK]E[K DR]AIK[ER]Q[EKN]AK[E]QJ [KR]D[RKE]R[KN]K[REN]R KTN[S]KJ[N]QT[S]D] (SEQ ID NO: 189)
DHR37 design	SSTERAAQSVKKYLQQQK DPDQAQKKAQEVKENIEKE ANS (SEQ ID NO: 196)	SSVIRAAA VVFFYLLEQGY DPDQALKKAQEVARNIENE ANS (SEQ ID NO: 197)	DDVIKEAAKVVKRLEEGQ DPDKALEEARKRAQKTEKK TTS (SEQ ID NO: 198)
DHR37 variant s	S[TD]S[AQ]T[SA]E[KQD]R[K E]AAQJ[RD]K[S]AE]VKJIRYK JERJYLQJ[K]Q[REK]QGGY RJD]SN]P[S]D]E]Q]E]AQ]ED K]K[R]K[VQ]AQ[RED]E[Q]V K[AQ]E[KT]N[QAD]E[T]K[E J]EQT]AN]T]S[T] (SEQ ID NO: 193)	S[DE]S[AE]V[A]IR[KES]AAA A[E]VVF[E]KJYLLER[NQ]QG Y[Q]D[S]P[A]D]E]Q]KER]AL K]ER]K]QVE]AQ]R]E]KR]V AR]KNQ]N]ADQ]IEN]KDE]E QT]ANS[T] (SEQ ID NO: 194)	D]E]D]ESQ]V[A]IK]RE]E]RT A]AAK]ERS]VVY]E]IK]K]E]R J]LE]E]KQR]E]RK]GQ]YKR] D]S]P]A]D]E]K]QDE]ALE]KQ J]E]KQR]AR]I]Q]K]ER]R]EHQ] AQ]KER]K]ENR]T]E]K]E]K]R EN]K]ETQ]T]E]KR]T]S]DT] (SEQ ID NO: 195)
DHR39 design	SDLQEVADRIVEQLKREGRS PEEARKEARRLIEEIKQSAG GD (SEQ ID NO: 202)	SELIEVAVRIVKELEEQGRSP SEAAKEAVELIERIRRAAGG D (SEQ ID NO: 203)	SDRIKKAVELVRELEERGRS PSEARRAVEEIQRSVEEDG GN (SEQ ID NO: 204)
DHR39 variant s	S[ND]D]E]KN]L]T]E]D]E]K]D]E] KNR]V]I]AD]K]E]R]K]EN]I]V] R]E]K]R]Q]AD]L]A]K]E]Q]R]R] KN]E]D]KN]GR]QHK]S]DN]P] ER]E]DN]E]S]AR]E]K]K]RE]E] [TK]Q]AR]DEK]R]E]K]L]I]V]E] KR]N]E]K]RQ]K]RQ]Q]DKE]S] [A]AGGD]NT] (SEQ ID NO: 200)	S[EDQ]E[DQ]N]L]I]R]E]RDQ]V [I]AV]A]I]R]QEW]IV]I]K]E]Q] E]QAD]L]A]E]Q]T]E]K]N]Q]Q] DKE]GR]QY]S]DN]P]A]S]AR] E]R]AAK]ER]E]TK]AV]A]E]R] J]I]V]E]KQR]R]K]IR]V]R]DE Q]AAGGD]N] (SEQ ID NO: 200)	S[DP]D]KE]R]L]IK]E]Q]R]K]R E]AV]A]I]E]K]L]I]ET]V]I]R]KE D]E]KQ]L]AE]E]QAN]E]KRN J]R]KED]GR]QKN]S]DN]P]A]S [AR]E]KDN]AAR]E]K]R]E]KQ] AV]A]E]RQ]E]RDK]I]V]Q]E]K A]R]KNE]S]DER]VE]RDK]E] KNR]D]N]Q]GGN (SEQ ID NO: 200)

	199)		NO: 201)
DHR40 _design	SESDEVAKRISKEAKKEGRS EEFVKELVERFREAIEKLKE QGD (SEQ ID NO: 208)	SEAIRVAVEIADEALREGLSP EEVVELVERFVQAIQKLQEN GE (SEQ ID NO: 209)	EDEIQKAVETAQEQLEEGRS PKEVVETVEEQVKEVEEQ QKGE (SEQ ID NO: 210)
DHR40 _variant s	S[TD]E[DKQ]S[A]D[EK]E[K] V[A]AK[QE]R[KN]S[A]EK[K] ER[E]QL[AKK]R[E]DKQ[GR] KAE]S[D]E[P]E[DK]E[QR]VK [NQ]E[K]LV[A]E[KR]R[D]F Y[R]KEQ[E]KQD[A]L[E]KQ D[K]ER[LK]QRE[E]KRD]Q[N ED]GD[N] (SEQ ID NO: 205)	S[EDK]E[DKR]AI[EKV]R[E]Q K[V]A[A]VE[RKQ]IAD[E]E[Q L]AL[Q]R[K]E[DK]GL[KRA]S [D]P[A]E[KQ]E[QRT]VV[A]E R[LV]E[KQ]R[E]F[Y]V[A]Q KRD]AI[L]Q[ENK]K[DQ]LQ RE[E]KOR]N[E]Q[GE]ND (SEQ ID NO: 206)	E[DKS]D[E]E[SAR]I[EKV]Q EK[K]RQ]AVE[RKQ]TIED]A Q[E]E[KNQ]Q[A]L[Q]E[RDK [E]DTK]GR[KAE]S[DN]P[A]K [ER]E[QKS]VV[A]E[RK]T[DR N]VE[Q]E[RK]Q[HES]V[A]K [ET]E[KNR]V[E]N[E]DQK[E] KR]K[EL]Q[DEK]Q[KRD]K[E QR]GE[QKN] (SEQ ID NO: 207)
DHR41 _design	SDIEKAKRIADRAIDVVRKA AEKEGGSPEKIREALQQAAR CAEKLIRLVKEAQESNS (SEQ ID NO: 214)	SDVREAAARVALEAVRVVVR AAEEKGCSPEEVVEAVCRA VRCAEKLIRLVKRAEESNS (SEQ ID NO: 215)	ENVRESARRALEKVLKTVO QAEFEKGSPPEEVVEQVCRS VRKAEQIRETQERERSTS (SEQ ID NO: 216)
DHR41 _variant s	S[DT]D[NLR]I[ARE]E[KDR]K [ER]AK[ER]R[KE]I[V]AD[KE Q]R[EK]AI[VE]D[EKR]V[A] V[A]R[QED]K[ER]AAE[KDR] K[NR]E[KQR]GGS[D]P[SE]E DNQ]K[ER]R[KDQ]E[QKR]A L[EIR]Q[KDE]Q[ERD]AK[RE] R[EK]C[AV]AE[KR]K[RL]L A[I]I[KLR]R[K]E[L]AV[V]A]K [EQ]E[KRQ]AQ[EDK]E[RDK] S[LAK]N[PS]S[N] (SEQ ID NO: 211)	S[DEQ]D[NA]V[A]R[QK]E[K RQ]AAR[EKQ]V[I]AL[I]E[RD Q]AVR[EK]V[A]V[A]VR[EK] AAE]Q[E]K[R]K[RET]GGS[D N]P[A]E[DKR]E[DQR]V[I]V A]E[R]AV[I]C[EAI]R[E]AV[A R]EK]C[AV]AE[RK]K[ERL]L [AI]I[VL]R[EKD]L[I]AV[V]A] K[EAQ]R[EDK]AE[Q]E[RDK] S[D]N[PSQ]S[N] (SEQ ID NO: 212)	E[DOT]N[RDE]V[A]R[E[KR]S AR]AR[KOE]R[KE]AL[I]E[K] K[HDE]VL[ER]K[ER]T[V]VQ [REK]Q[KEN]AE[QKS]E[KR] E[DKR]GK[G]S[D]P[A]E[DR K]E[KD]V[I]V[A]E[R]Q[RND] V[I]C[EKQ]R[EK]S[A]V[A]R EK]K[QRAE]AKQ[E]KRO]Q EDR]I[VL]R[KEQ]E[KTD]T[Q VA]Q[EAK]E[KNT]R[KE]E]Q [R[KDE]S[RK]T[NPS]S[DN] (SEQ ID NO: 213)
DHR42 _design	SDAEVKKQAEIINRAYK TAQKQGESDSRAKKAELV RKAAEKLARLIERAQKEGD (SEQ ID NO: 220)	SDALEVARQALEIARRAFET AKKQGHSAEAAKAFVDV VEAAISLAELIISAKRQGD (SEQ ID NO: 221)	QKALEIARKALQKAKENFE EAQKRGESATQAAKRFVDT VEKEIKKAQEQIKRERKGD (SEQ ID NO: 222)
DHR42 _variant s	S[DT]D[TIQ]A[S]E[KQ]E[KR Q]V[I]K[ERQ]K[ED]Q[EDK]A E[KR]E[K]I[LT]AN[EQK]R]Q KE]AY[EKR]K[EDR]T[QER] AQ[KRE]K[EQR]Q[DE]GE[Q HK]S[D]D[EPS]S[DK]R[E]Q[A K]QDE]K[Q]AE[YFR]K[EDR] L]TDA]VR[EKL]K[RE]AAE[R KD]K[EQR]LAR[EKQ]L[A]E [KR]R[KE]AQ[ERK]K[DER]E [QN]GD[NS] (SEQ ID NO: 217)	S[DEQ]D[TEI]AL[A]EK[E]KQ R]V[I]AR[E]Q[E]AL[A]E[K]I LT]AR[KE]I[R]KDE]AFE[KR] T]EQ]AK[RNT]K[R]Q[DER]G H]QLE]SAT]QR]E[QR]AAK[E]AF[Y]V]AEQ]D[TAL]VVE[R KD]AAI]K]S[KEQ]LAE[QRT] L]I]A]I]EL]S[KEQ]AK]QRE]R KQ]Q[ED]GD[NS] (SEQ ID NO: 218)	Q[DER]K[EDT]AL[E]K[E]KQR I]V]AR]ES]K[EQ]AL[A]Q[EK R]K[RD]AK[ELR]E[RK]N]AE]FE[KQR]E[QKN]AQ[REN]K NR]R[DKQ]GE[KLR]S[D]AT EQR]Q[ER]AAK[QE]R[EA]F Y]V]AEK]D[ER]T[VR]VE[KD R]K[E]E]A]I]REK]K[E]K[E]A Q[ERK]E[KR]Q[ASE]I[RLN]K [ERQ]R]LE]E[QDK]R[KEQ]K ER]GD[QKT] (SEQ ID NO: 219)
DHR43 _design	SKEEELIEKARRVAKEAIEE AKRQKDPSEAKKAAEKL KAVEEAVKEAKRLKEEGN (SEQ ID NO: 226)	SELAEISEAIQVAVEAVEE AVRQKDPFKAAEAAAELI RAVVEAVKEAERLKREGN (SEQ ID NO: 227)	SELAKKINDTIREAVREVQ AVEDGKDPFEAAAREAAEKI RESVERVREEEKKRRGN (SEQ ID NO: 228)
DHR43 _variant s	S[TD]K[ETD]E[L]E[KD]E[KN Q]LIE[KR]K[ER]AR[E]R]EKQ [V]AT]AK[ER]E[KRN]A[L]I V]E[KDR]E[KQT]AK]QRE]R KED]Q[DK]GK]QL]D[NS]P[E S]S[DN]T]E[KRL]AK]REQ]K ER]AAE[KDR]K[ER]LIK]ENR]AVE]R]E[KQR]AV[A]K]E]E TVK]AK]ER]R]KE]LK]ER]E	S[EQ]T]E[DK]LAE[RKD]LIS[E KR]E[KR]AI]Q]REK]V[AT]AV [I]E[RDQ]A[L]VE[DKQ]E[QT R]AV]QRA]R[KE]Q[DE]GK]Q L]D]N]P[A]F]WA]K]RED]AA E]K[R]AAAE[RK]LIR]KE]AV V[A]E[KRD]AV[A]K]ERQ]E VR]AE[RH]R[KQ]LK]ES]R]E K]E]NDK]GN (SEQ ID NO:	S[KET]E[DK]LAK[RDE]K[ER]N]KRE]D[EQQ]T]AS]IR]EK Q]E[QK]AV[IL]R]KEQ]E[DN K]V[I]Q]E]Q[END]AV]QAN]E [KR]D]QKE]GK]Q[D]NT]P[A] F]WAT]E]EDK]AAR]EKH]E]R KD]AAE[KQR]K[ERH]IR]EK Q]E[KNQ]S[VET]V[A]E[KRD R]QED]V]A]R]QKE]E[KR]E]

	RKD[E]NQRJGN (SEQ ID NO: 223)	224)	DKQ[E]AS[E]KR[K]RA[K]DR[R]KEN[R]NEKJGN[KEQ] (SEQ ID NO: 225)
DHR44 _design	SNEQEKDLKKAEEAAKSPDP ELIREAIERAEESGS (SEQ ID NO: 232)	NKAKEIILRAAEFAAKSPDP ELIRLAIEAAERSGS (SEQ ID NO: 233)	EKAKEIILRAAEFAAKSPDP ELQKLAKAEARERLG (SEQ ID NO: 234)
DHR44 _variant s	S[T]N[DT]E[DE]Q[DE]E[KDN] J[K]EQ[R]K[ER]D[R]K[LK]ER D[K]RDE[AE]KQR[E]KQ[AA] K[ENR]SP[ST]D[N]PE[DNS]L [KND]I[L]R[KDE]E[RKT]AI[L] V[E]KDR[R]ELQ[AE]QKD[E] KRQ[S]QET[GS]T (SEQ ID NO: 229)	N[ED]K[RE]Q[AK]E[E]K[I]LR DEI[AAE]KR[E]V[AAK]DER S[A]P[ST]D[N]PE[DNQ]LI[L] R[KDE]L[TKQ]AI[V]E[KR]A W[AE]KQR[R]E[S]Q[GS]T (SEQ ID NO: 230)	E[D]K[DE]Q[AK]R[E]K[R]IK REL[R]I[L]T[AAE]DKR[E]VO AQ[KE]K[R]N[S]AE[P]SQ[D] N[P]E[E]DN[L]Q[L]K[ER]L[KET] [AK]EQ[R]E[KR]A[W]R[AEK] E[KRN]R[E]KQ[L]QSEJG (SEQ ID NO: 231)
DHR45 _design	SSEEELEKDAEASESGAD PEWLREIVDLARESGD (SEQ ID NO: 238)	SEVIELAKRALEAAKSGADP EWLLRIVRQAEESGS (SEQ ID NO: 239)	EEVIELAKRALEAAKSGADP KELLEVRKREESG (SEQ ID NO: 240)
DHR45 _variant s	S[DN]S[DT]E[D]E[TS]D[E]KD E[KR]L[E]Q[K]R[D]LKA[AR] [KD]E[S]AS[A]E[NK]S[T]GA D[TN]P[S]E[NT]W[ALY]LR[K] E[E]KR[IVD]REN[L]QTD[AR] KTS[E]KNR[S]Q[GD]NT (SEQ ID NO: 235)	S[DP]E[Q]V[A]I[K]E[K]LAK ES[R]LAK[AL]E[QD]AAK[E]S [T]GAD[NT]P[A]E[RKQ]W[A] LY[LL]W[R]KQ[VR]QDE[Q] TE[AE]RST[E]KDN[S]E[GS]D N (SEQ ID NO: 236)	E[PD]K[E]QND[V]A[I]K[E]KR L[EA]AKR[E]KQ[AL]E[DRK]E [RD]AK[R]K[E]GK[QH]D[NT] P[A]K[RE]H[E]QDR[LL]W[E] KR[E]K[VR]QKA[K]ET[R]KN S[E]T[E]KDR[S]RKEJG (SEQ ID NO: 237)
DHR46 _design	STKEEKERIERIEKEVRSPDP ENIREAVRKAEEELLRENPS (SEQ ID NO: 244)	TEAEELLRRAIEAAVRAPDP EAIREAVRAEEELLRENPS (SEQ ID NO: 245)	EEAKELLRRAIESAKKAPDP EAQREAKRAEEELRKEDP (SEQ ID NO: 246)
DHR46 _variant s	S[D]T[D]K[DE]Q[E]DKT[E]KL Q[K]RED[E]KDR[R]TDK[I]E AK[E]KR[R]EKD[IE]RDK[K] R[E]A[V]A[R]EKD[S]A[P]S[D] N[P]ADS[E]DKN[N]EAD[IR] EK[E]KQR[AVR]EK[K]EAD AE[ARK]E[KR]LL[YA]R[KE] Q[E]KRN[N]D[P]D[S] (SEQ ID NO: 241)	T[DE]E[D]AE[KQR]E[KR]L[A] I[LR]EAS[R]KE[A]E[RKQ]A RQ[AV]A[R]EKD[APD]N[P]A D[E]SDK[AIR]KE[E]ALR[AV] R[ED]AAE[RS]E[QH]LL[Y]R EK[E]NRD[N]D[P]D[S] (SEQ ID NO: 242)	E[DQ]E[DK]AK[QER]E[KR]L [AI]LR[EK]R[ETK]A[E]RQ[Q] S[AER]AK[QE]K[ERN]APD[N] P[SEK]E[DKN]AQ[R]R[KDE] E[AKL]AK[E]E[EQK]AE[QK] R[E]KR[E]QDR[LR]DKE[K]E R[E]NQD[D]N[P]D[S] (SEQ ID NO: 243)
DHR48 _design	NSREEEEAKRIVKEAKKSGF DPREEVEKALREVIRVAEETG N (SEQ ID NO: 250)	SEALKEALKIVEEAAKSGYD PAEVAKALAEVIRVAEETG N (SEQ ID NO: 251)	PEELKEALKRVLEAAKRGE DPAQVAKELAEIIRRNQEEG (SEQ ID NO: 252)
DHR48 _variant s	N[D]S[D]R[EDH]E[AS]E[KR] E[K]E[KDL]AK[ER]R[EK]I[V] V[A]K[E]E[QRK]AK[Q]K[E]S GF[Y]D[N]P[S]E[NTK]E[KQT] [VE]KQ[K]RE[ALR]DEK[E]R KQ[V]L[I]QR[R]EK[V]AE[QR] E[RQ]T[KH]GN[D] (SEQ ID NO: 247)	S[PR]E[D]AL[A]K[ER]E[DKQ] [ALK]RED[I]V[V]A[E]KR[E] Q[AAK]ER[SGYD]N[P]A[AE] QD[VAK]RD[ALAE]KR[V]L[I] R[KE]VAE[Q]E[DR]T[HKS]G N[D] (SEQ ID NO: 248)	P[RQ]E[D]E[SAD]L[A]K[ENR] E[KR]ALK[ER]R[E]V[A]L[E] RS[E]KR[AAK]ER[R]KEQ[GE] [KRT]D[N]P[A]AQ[DKE]VAK [ED]E[K]LAE[KR]E[Q]IR[EK] Q[R]EDK[N]ARD[Q]ET[E]RD K[E]KRJG (SEQ ID NO: 249)
DHR49 _design	DSEEEQERIRRLKEARKSGT EESLRQAIEDVAQLAKKSQD (SEQ ID NO: 256)	SEVLEEAIRVILRIAKESGSE EALRQAIRAVAEIAKEAQD (SEQ ID NO: 257)	PRVLEEAIRVIRQIAEESGSE EARRQAERAEIIRRAQ (SEQ ID NO: 258)
DHR49 _variant s	D[TS]S[T]E[D]E[DE]Q[E]WS[Q] KAE[E]KNR[R]KDN[I]A[R]K E[R]KEN[I]T[L]AVVW[K]EN R[E]KND[AR]QTD[K]NR[S]D Q[GT]SDN[E]DKN[E]DS[S]A DQ[L]I[R]KIQ[Q]KER[A]E[K] DN[D]KRE[V]I[AQ]RKE[L]IE V[AK]ERS[K]EDQ[S]A[Q]TN R[D]TS (SEQ ID NO: 253)	S[P]E[DS]VL[W]E[KAR]E[RH] L[A]I[A]R[EDK]V[ER]L[I]L[AE] V[R]EK[I]AL[AK]DEQ[E]QD R[S]A[GS]DN[E]DNP[E]DR[A] [V]L[I]R[KAI]Q[ERK]AIR[ED] Q[AV]I[AE]RDK[I]AK[ERS]E QDK[AQ]TND[D]NST (SEQ ID NO: 254)	PR[NED]VL[W]E[KR]E[TAH] AI[KAQ]R[KE]V[ERQ]IR[QE] K[Q]ERK[I]AL[AE]KDR[E]Q ND[S]A[GS]DN[E]DPS[E]DK A[V]R[KE]R[KE]Q[KEL]AE KQR[R]EK[AE]IKQ[E]RDK[E] [RQT]IR[KDE]R[KD]R[QDK] AQ[TND] (SEQ ID NO: 255)

DHR50 _design	DPEEVRREVERATEEYRKNP GSEAREQLKEAVERAEEA ARSPD (SEQ ID NO: 262)	PEAVQVAVEAATQIYENTP GSEAKKALEIAVRAAENA ARLPD (SEQ ID NO: 263)	PEAVRVAEEAADQIRKNT GSELAKRADIEKKRARELLE RLP (SEQ ID NO: 264)
DHR50 _variant s	D[SNT]P[ST]E[DS]E[DR]V[A EL]R[KEL]R[KDE]E[TKI]V[A E]R[KD]R[KED]AT[EKQ]E[K R]E[IRT]Y[W]R[KQD]K[E]N[HRD]PGSD[ER]E[DK]AR[KE H]E[KR]Q[AS]L[K]R[E]E[KRQ] AVE[K]R[AD]AE[KQR]E[K]A L]AR[KDE]S[LKN]P[S]D[SN T] (SEQ ID NO: 259)	P[WAE]E[KD]AV[AL]Q[EDK] V[AT]AV[A]E[RKN]A[I]AT[K QE]Q[IKR]I[V]Y[W]E[QDK]N [DT]T[E]PGSE[DO]E[L]AN[A K]E[R]K[ERT]ALE[KR]I[L]A[A VR]DE]AAE[R]N[AER]A[L]A R[EN]L[SN]P[S]D[SN]T (SEQ ID NO: 260)	P[ST]E[KDQ]AV[AL]R[KED] V[TAK]AE[QKR]E[RKT]A[I] AD[KE]Q[IRT]I[V]R[WY]K[QE]N[T]T[E]PGSE[D]L[AEN] AK[E]R[EKT]AD[QKR]E[KN R]I[L]A[K]A[K]E[R]E[KQ]AR[EQK]E[KQR]L[NVA]L[KAR] E[KDN]R[KE]L[SAN]P[S] (SEQ ID NO: 261)
DHR51 _design	QSEDRKEKIRELERKARENT GSEARQAVKEIARIAKEAL EEGN (SEQ ID NO: 268)	ADTAKEAIQRLEDLARDYS GSDVASLAVKAIKIAETAL RNGY (SEQ ID NO: 269)	KETAEAAIKRLRELAEDYK SEVAKLAEAAIERIEKVSRR G (SEQ ID NO: 270)
DHR51 _variant s	Q[HNK]S[DNT]E[D]D[EQT]R [QAD]K[IQ]E[KR]K[DR]R[K QE]E[R]L[E]AQK[R]K[E]K[TI] AR[EKQ]E[KRT]N[YEH]T[S] GS[T]D[E]E[DKR]AR[K]Q[KE N]AV[A]I[K]QER[E]DKR[AR] KED[IAK]ER[E]KQ[AL]A[E] KRQ[E]R[K]GN[S] (SEQ ID NO: 265)	A[SRK]D[E]T[V]AK[E]Q[E]H T[A]Q[E]KR[R]E[L]E[AQK]D[K]L[I]V[AR]E[DS]D[KTE]Y[F] S[T]GS[T]D[ES]V[A]AS[RK]L [E]AV[A]I[K]E[RQ]A[L]IAK[E HR]I[A]E[KQ]T[VER]AL[A]R[KEQ]N[Q]GY[ND] (SEQ ID NO: 266)	K[RST]E[D]T[V]AE[KQ]E[HT Q]A[I]K[K]RD[R]EQS[L]R[QK E]E[KDS]L[I]V[A]R[E]DKR[D] KQE]Y[F]K[TED]GS[T]E[DO] V[A]AK[E]L[EDQ]AE[KR]E[KQ]A[L]I[E]KQR[R]EHK[E]K N[K]EDT[V]EIQ[S]A[R]K[E]K R[R]QEN]G (SEQ ID NO: 267)
DHR53 _design	SNDEKEKLKELLKRAEELA KSPDPEDLKEAVRLAEVV RERPGS (SEQ ID NO: 274)	NLAKKALEIILRAAEELAKL PDPEALKEAVKAAEKVVRE QPGS (SEQ ID NO: 275)	ELAKKALEIIERAAEELKKSP DPEAQKEAKKAEQK VREER PG (SEQ ID NO: 276)
DHR53 _variant s	SN[DS]T[D]E[E]DTR[K]ED[E] K[K]E[LK]ER[E]K[L]IRK[LK] DRE]R[K]AE[R]E[KQ]LAK[R EN]S[L]PD[N]P[E]E[NDK]D[A]LK[ERQ]E[KR]AVR[DEK] L[TAE]AE[KQ]E[KR]VV[I]R[EQK]E[Q]R[Q]P[S]GS (SEQ ID NO: 271)	N[EDS]L[NAK]AK[E]K[ERT] ALE[KR]IILR[TKD]AAE[KR] E[AN]LAK[EQR]LPD[NS]P[E] E[NRT]ALK[EQN]E[KRA]AV K[ER]AAE[K]K[DEQ]VV[I]R[NDK]E[TQ]Q[RT]PGS (SEQ ID NO: 272)	E[SR]L[ANQ]AK[E]K[ER]AL E[KDR]IIE[QK]R[ELT]AAE[K RT]E[NA]LK[QER]K[R]S[L]P D[NS]P[DE]E[QKN]AQ[KR]K [E]E[KRD]AK[E]K[RE]AE[K NQ]Q[EKN]K[ERD]VR[K]E[D RK]E[TQ]R[QN]PG (SEQ ID NO: 273)
DHR54 _design	TTEDERRELEKVARKAIEAA REGNTDREVLQRALEIAR ESGT (SEQ ID NO: 280)	TEAVKLAEVVARVAIEAA RRGNIDA VREALEVALEIA RESGT (SEQ ID NO: 281)	EEAVRLAEVVKRVSDEAK KQGNEDAVKEAEVVRKKIE EESG (SEQ ID NO: 282)
DHR54 _variant s	T[S]T[DNS]E[DO]DE[WAD]R [KEN]R[E]K[E]KQN[L]I[E]K[R Q]K[E]V[A]I[A]K[R]E[KQ]K[R E]AI[K]A[E]KQR[AAR]QEK[E [KRD]GN[D]T[ANR]D[E]K[E] RQ]VR[EKQ]E[RK]Q[A]LQ[E RD]R[KEN]A[I]L[AR]E[KQR I]AET[AR]NKE[E]KDO[S]A QT[GT]S (SEQ ID NO: 277)	T[DEK]E[KD]AV[WF]K[EDR] L[RK]AL[I]E[RD]V[A]V[A]A [K]R[E]KQ]V[A]AI[K]AQ[E]A QR[AAR]Q[R]QK]GNT[ANR] D[E]AVR[E]K[E]AIK]ALE[RK I]V[A]A[I]L[A]Q[E]Q[RK]I[A]A R[KND]E[DNK]S[A]GT[S] (SEQ ID NO: 278)	E[DKP]E[RD]K[AV]WF[R]KD E[L]W[R]E[AL]EIQ[E]K[R]V[A] V[A]I[K]DEH[R]EKN]V[A]S[A D]EKR[E]AKQ]AK[QED]K[E NR]Q[RK]GNE[SRD]D[ES]A V[AES]K[EDQ]E[LRA]AE[QR K]E[KQ]V[AT]R[A]E[K]ER] K[REH]I[A]E[KQR]E[KRN]E NDK]S[AK]G (SEQ ID NO: 279)
DHR55 _design	SSVAEEIEKRAKKISKELKK EGKNPEWIEELQRAADKL V EVARRATS (SEQ ID NO: 286)	SDALEIAKRAVKIAEELAKQ GSNPKWIAELLKAAAKLVE VAARATS (SEQ ID NO: 287)	PKALKQAKEAVKEAEELAK KGRNPKEIAEELKKRAKEVE KLARST (SEQ ID NO: 288)
DHR55 _variant s	S[DN]S[NT]V[K]AE[DK]E[KT I]E[RKA]K[E]R[TIK]AK[E]K[E]IS[A]K[ER]E[H]LKK[R]E[Q]GK[AS]N[D]PE[TN]W[AK]IE [QNK]E[KNR]LQ[L]R[DEK]A AD[ENK]K[ER]L[V]A[E]R[K]V [A]AR[QK]R[KEN]AT[Q]S[N T] (SEQ ID NO: 283)	S[PE]D[V]K[AL]R[E]K[IAK]E QL[R]KLT[AV]A[K]EQR[IAE [KL]R[E]RDK]LAK[ERQ]Q[ER]GS[A]N[D]PK[ES]W[AKQ]IA E[K]LLK[EQ]AAK[EDQ]LV [A]E[RQ]V[A]AA[D]R[E]K[AT N]S[N] (SEQ ID NO: 284)	P[SE]K[ED]AL[RW]K[R]E[Q] TE[AK]EQ[E]KRD[AV]A[K]E R[E]KQS]AE[K]E[KR]L[RKD] AK[DR]E[K]R[E]GR]QDK[N]D] PK[ES]E[K]IAE[K]E[RHK]LK [ER]K[E]R[A]AK[ER]E[KTR] V[A]E[RKL]K[EN]L[QA]A[D] R[KE]S[EQR]T[Q] (SEQ ID NO: 285)

DHR57 _design	STEELKKVLERVRELSEKAK ESTDPEEALKIAKEVIELALK AVKEDPS (SEQ ID NO: 292)	TDALRAVLEAVRLASEVAK RVTDPKALKIAKLVIELAL EAVKEDPS (SEQ ID NO: 293)	EEAKRAVEEAKRLAEVSK RVTDPELSEKIRQLVKELEE EAQKEDP (SEQ ID NO: 294)
DHR57 _variant s	S[D]T[DN]E[D]E[DK]L[K]ER K[Q]VL[KIY]E[RK]R[DTK]V R[EKQ]E[HR]L[AD]S[A]E[KR D]R[EQ]AK[REN]E[K]S[VIE] T[DS]D[N]P[T]E[DTN]E[KDN]AL[KAR]K[E]L[V]AK[E]E KR[V]A][V]E[KR]L[E]AL[AE K]K[EQ]AV[A]K[RDE]E[K]D NK]PS (SEQ ID NO: 289)	T[DE]D[LNT]ALR[EKQ]AVL YAE[E]RKL[AVR]EKQ[LAS] A[E]R[V]A]AK[QER]R[K]V[I L]T[N]D[N]PD[E]K[AL]AL[A KR]K[E]L[V]AK[ER]L[KW]V [A]I[V]E[KR]LAL[EA]K[E]KD L]AV[A]K[ENR]E[QNR]D[N] PS (SEQ ID NO: 290)	E[DKQ]E[LDT]AK[L]A[R]EK AVE[K]E[L]RK]AK[IEA]R[ED K]LAE[KQR]E[RKQ]V[A]S[A]K[QER]R[NOK]V[L]T[D]DP [DS]E[NDK]L[KNS]S[AR]E[K RN]K[ER]I[L]V[R]K[E]Q[ERK] L[EKW]V[AK]K[ER]E[KRD]L E[RKQ]E[KR]E[L]RD]AQ[KE N]K[ER]E[QRH]D[NY]P (SEQ ID NO: 291)
DHR59 _design	KTEVEKKAKEVIKEAKELA KELDSEEAKKVVERIKEAAE AAKRAAEQ GK (SEQ ID NO: 298)	TEVAKLALKVLEEAIELAKE NRSEEALKVLEIARAALAA AQAAEEGK (SEQ ID NO: 299)	SDEARDALRRLEEAIEEAAKE NRSKESLEKVRREEAKEAEQ QAEDAREG (SEQ ID NO: 300)
DHR59 _variant s	K[N]T[S]E[K]T]VE[KDR]K[ED]K[QET]AK[ER]E[KR]V[A]I KR]K[E]E[KRN]AK[ED]E[KR N]L[V]A]REV[K]RE[E]DKN L]IA]D]KPR]SE]DKQ]E]TVL AK[E]K[DQR]V[V]A]E[K]R]E AQ]I[V]K]R]E[KR]AAE[K]A] E]AK[R]IE]R[EKQ]AAE[KDR] Q[SN]GK[N] (SEQ ID NO: 295)	T[S]E[R]VAK[E]L[RK]ALK[E]V[A]LE[K]T]E[RQ]AIE[RK]L V]AK[EQR]E[KN]N[L]A]R]D KP]SE[KD]E]TKQ]ALK[E]VV]A]L]A]E]AQ]I[V]A]R]KE]AA L]KAE]A]E]AAQ]ER]AAE[K RQ]E]QSD]GK[N] (SEQ ID NO: 296)	S[T]D[ER]E[V]AR[KE]D]KER]ALR[EKD]R]KE]LE[KQT]E KQR]AIE[RK]E]HTD]AK[EQ R]E]KQR]N[DHR]R]DKP]SK DE]E]D]S]A]LE]KNQ]K]E]V A]R]LKY]E]DK]E]RIW]AK[R E]E]KQN]AE]KRA]Q[EK]Q[E KD]AE]RD]D[RKN]AR]KQS E]NR]G (SEQ ID NO: 297)
DHR60 _design	TDIKKKAEEIIEAKKQSGE DAIRLAQEAKKQGT (SEQ ID NO: 304)	DILVRAAEIVVRAQEQGSED AIRLAKEASREGT (SEQ ID NO: 305)	PTLVKAAEKVVRAQQKGSQ DTIEKAKEESREG (SEQ ID NO: 306)
DHR60 _variant s	T[ND]D[TS]IIT]K]SQ[R]K]DE K[ED]AE[KDN]E[RK]I[V]A]I K]KE]RD]AK]QE]KQ]TEN]G SE]DRS]D]TEK]A]I[K]R]EK]L]AT]AQ]ERK]E]RK]AK]A]K ERN]Q]KRE]GT]ND] (SEQ ID NO: 301)	D[EPK]IIT]L]A]V]A]R]KDE]A AE]RKQ]I[V]A]V]I[V]A]R]E] AQE]QKR]Q]EST]GSE]DRS] D]TA]AIR]EK]L]AT]AK]REA]E]KQR]AS]A]R]E]E]QRK]GT [ND] (SEQ ID NO: 302)	P[EQR]T]RIK]L]A]V]A]K]ER] AAE]QOR]K]RE]V]I[V]A]R] EDK]AQ]ET]Q]KRE]K]E]GSQ [DER]D]E]T]SK]I]K]E]KR]K] RQ]AK]REN]E]KR]E]ADK]S] A]R]KE]E]KRQ]G (SEQ ID NO: 303)
DHR62 _design	DNDEKRKRAEKALQRAQEA EKKGDVEEAVRAAQEAVR AAKESGD (SEQ ID NO: 310)	NDVLRKVAEQALRIAKEAE KQGNVEVAVKAARVAVEA AKQAGD (SEQ ID NO: 311)	QDVLRKVSEQAERISKEAK KQGNSEVSEARKVADEAK KQTG (SEQ ID NO: 312)
DHR62 _variant s	D[SN]N[T]D[ER]E[D]K]L]R]K E]K[EQ]R]EK]AE[KRQ]K[ER D]AL]I]Q]KER]R]EKN]AQ[K ED]E[KQ]AE[RQ]I]K]R]K[ED R]GD[N]V]A]E[KDR]E]RSK] AV[A]R]KE]AA[L]Q]EK]E]R] AV[A]R]EKQ]AAK[TR]E[KR] S[A]GD[SN] (SEQ ID NO: 307)	N]QKT]D]E]V]LA]LR]KE]K]E QR]V]A]AE[R]Q]VEA]AL]E] R]KQ]I]AV]AK[EQD]E]QL]A E]RQ]K]RE]Q]ED]GN]D]V]A] E]KRQ]V]AL]AV]A]K]EDR]A A]L]R]KE]V]I]AV]A]E]RD]A AK[SRE]Q]EN]AGD[S] (SEQ ID NO: 308)	Q[KT]D]ES]V]LAS]LR]EHK] K]ER]V]A]S]A]E]RQ]Q]VAE] AE]KQR]R]KEQ]I]AV]S]A]K] E]E]QDL]AK]ER]K]R]Q]ED] GN]D]S]EKD]E]DQ]V]AL]S] A]E]KRD]E]KQ]AR]KQE]K]E Q]V]I]AD]KNR]E]KRH]AK]A L]K]RT]Q]NE]T]A]G (SEQ ID NO: 309)
DHR63 _design	DPDEDRLKEELKKIREAL REAKEKPDPEEIKRALREVL EAIIRILKLAERAGD (SEQ ID NO: 316)	PDLAREALKEINKVIREALEI AKRVDPDEVIKEALRVLEA IRAILKLAERAGD (SEQ ID NO: 317)	PDLAREALEEDKVIDEAQEI SERVPDEEVQREAEQVIKEA DRARKKLSEQSG (SEQ ID NO: 318)
DHR63 _variant s	D[N]P[SNT]D[E]E[DK]D[A]R] EAK]E[KR]R]DEK]L]A]K]ER Q]E[KRQ]E[AV]L]A]V]K]ER] K[DIL]I]A]R]EK]E]KQR]AL]I AS]R]KE]E[KDI]AK[REN]E]K T]K]I]VT]PD[N]P]ST]E]NDQ]E]QTD]IK]ALR]R]EK]ALR]EK	PD]NEK]LAR]KEA]E]KHR]A [VI]L]A]K]ERD]E]AK]I]AV]N [ALE]K]RE]V]AL]I]A]R]KEQ] E]DIN]AL]AI]Q]E[KR]I]A]AK] ETQ]R]KTE]VPD[N]P]T]E]NT K]VIK]AER]E]AKT]ALR]EK N]VV]I]AL]AKQ]E]TAQ]AI]L	PD]ENK]LAR]KE]E]KR]A]VI] L]AR]E]K]RE]AQK]I]AV]D]K E]K]ER]V]AL]I]AR]D]KER]E] IVN]AQ]EKS]E[RKN]I]A]S]A KE]E]KNR]R]EKT]VPD[N]E] PS]E]NKT]VQR]KEQ]E]QA]A Q]KRE]E]KR]VI]AV]K]EDR]

	Q E I K V I A L K A Q E K R A I L V R E K D R K D I A L R A K E Q L A E K Q R R K D Q A G D N (SEQ ID NO: 313)	V R Q E K A I A L A R K E Q D L A E K Q H A G D N (SEQ ID NO: 314)	E Q I K A D K Q E R K E Q A R A K I K E T K E R L S A E K E K Q Q H S A J G (SEQ ID NO: 315)
DHR64 design	DPEDELKRVEKLVEAEELL RQAKEKGSEEDLEKALRTA EEAAREAKKVLQAEKEGD (SEQ ID NO: 322)	PEVALRAVELVVRVAELL RIAKESGSEELERALRVAE EAARLAKRVLELAEKQGD (SEQ ID NO: 323)	PEVARRAVELVKRVAELLE RIARESGSEEAKEAERVRE EARELQERVKELREREG (SEQ ID NO: 324)
DHR64 variant s	D S P S E D K D E K T L V K E R R K V A E K R K E L I T E V K RED E K Q T A E D K Q E K Q L K A D L R R K Q E Q E K D A K Q N E K R K E G S E D K E D D A E L E K D R K E A L A E R R E K Q T R V A E A H N E Q R K A A R K E N E R A K E R K E V A L I A E K D H Q E K S A E K Q K E R E Q N D G D S (SEQ ID NO: 319)	P A E Q V A A L V R K E A V A E R L V R E V A A E K R L I L L R E K I A A K Q E N E Q D S S K R E G S E D R E D A E K Q T R E K A L A E R E K Q V A E S E K A A R K L Q A K E Q R D E V A L I A E D K L A E Q R K K R Q Q R G D (SEQ ID NO: 320)	P A E K R D V A A R K E H R E K T A V A E K R L V K Q R R E K V A A E K D R L I L E K R R K E Q I A A R K E N E Q D S S E K Q G S E D K A E K R K A E K Q R K Q V R E K Q E K D R E K Q R A R K E E K R L E Q E K R E K N R E Q V A K E D E K R L R A E K R R K E Q G (SEQ ID NO: 321)
DHR66 design	TSDDDKVREAEERVREAIER IQRALKKRDTDPARKALEA AKLLKVVVEKAKKRG (SEQ ID NO: 328)	SDAIKVAEAAARVAEAIARI LEALNERDTPDARKALRAAI KLAEVVYKAAESGT (SEQ ID NO: 329)	TEALKVAEKAARVAEKIARI LEKLNERTDTPARKKLRQAI KEAEKVYKESEQG (SEQ ID NO: 330)
DHR66 variant s	T S D N T D N E R D E Q D K E K R I V L R K E D E A E K R E K D Q R V A R E D E K Q R A I E Q E K R E K I A Q K R R E K Q A L I K K E D N R N K S D P T S D P D E S D E S A L R Q K K R E N A L E K A A I K Q E R K R L L A K R K E R V V I E K D K E R D A K E S Q K R E R E Q K G T (SEQ ID NO: 325)	S D T E D N R E A I A L K R V L A E A A R V A A E Q A I A A R E K I A L E A I N E K D E N K R R N K I D T D N P D D E S A L R K E D R A L V R K A A I I V K E L L A E D K V V I V A K E Q R A A E Q R D S R Q D G T (SEQ ID NO: 326)	T D R E D R S A L I A K E Q R V L S A E K K E Q A R K D V A A E K E S T I A A R D E I A L E D K K E R L I N K E R E K D R R N D I N T S P D E E D A E L R L K Q K E N L V R E K Q Q D R A I I V K R E R D K A E R K E Q V I V V K Q R E K S L S A E E K Q R Q K R G (SEQ ID NO: 327)
DHR67 design	TSEIDKLIKLRQTAKVVR EAEERKRRSTDPTVREVIER LAQLALDVAAEAARLIKKA TT (SEQ ID NO: 334)	SEVAKLVWKLARTAEVIRE AIERAERSTDPEVIRVILELA RLAAEVAKAEARLIVKATT (SEQ ID NO: 335)	EEVAKKVWKEAYRAIEIR KAIEKAERSTDPEIKKILEE ARKKAEEAIERAKEIVKST (SEQ ID NO: 336)
DHR67 variant s	T N D S T D E D R T I L K D K E K E L I V K K E R K R D E L V R Q E K Q K N R T E K Q A K D E K Q R V A K I A E R K E N E R D Q A E K E K R R A L K I Q R R K R K E N S A T D N S P I S D T R D N V L R A E K N T V I L E K H Q R E K L A I A Q E K R L I A L K R E D K R V I A E K D R E R N A A R K E Q L I K Q R E K A T E P Q T N D (SEQ ID NO: 331)	S K E T E D T K V L I A K E R L V I W A K E Q R L V A R A K N T E R A I L E K R D V A I V R A E K E D R A I A E R K Q R A L A V I E L Q A R E K N S A T D N S P D S E E T D N V L I A R K E V I L I L W E K R L A I A R K E L A A E K R S V I A K I Q E E R H K A A R E K L I V A K A T K P T D N (SEQ ID NO: 332)	E K R T E D N V L I A K E Q R K R V I W A K E Q R E L K T A Y K A Q R E K D A I L E K D E R K N I V R A E L K E Q R A I A E K D K D E Q A V I E R A K R K E S A E T T N Q D N S P T E Q N E T D E K D N I A K E R K E I E L W E K R E K N R A R K E K E K I E A A E K R E K R A I K E E K R R K E T A K E R E K R Q I Q T V A K N S D K T P (SEQ ID NO: 333)
DHR68 design	TPRERLEEAKERVEEIRELID KARKLQEQGNKEEAKEKVL EAREQIREVTRELEEIAKNS DT (SEQ ID NO: 340)	PELALRAAELLVRLIKLLIEI AKLLQEQGNKEEAKEKVL ATELIKRVTELLEKIAKNSD T (SEQ ID NO: 341)	PELAKRAAELLKRLIELLKEI AKLLEEGNEDEAEKVKKE AKELEERVRELEERIRKNSD (SEQ ID NO: 342)
DHR68 variant s	T P S T N R E K E D R K D Q L V E R K E K R A K E R E K Q R R K K V E K D Q E K I V R E K E K R L D K T I V D R K E K A R E K L Q A E L E K R Q S K N G N K W P S E D K T E K A E R K K E V A E Q L R E D E K Q R	P A V N E L I A L V R A A E K L L I V R K D E L I V K E R L L I V E R I A K E L L Q A L E R K N I Q S I G N K S T E D E D A E K R D K R D S V A L R E D K E K R T A T E R E K L Q A E I K E R R E K N V A I A E R E K R Q L	P A T V E L I A K Q E R A A E D K R L L I K E Q R E K L I V E K R L L K E Q R E R K T I A K E L L E K E R Q E S O N G N E S K P D E E D A E K N Q K R D E V A K E Q E K R D E K R D A K E N D E K Q L A D Q E K E K

	JAR[E]E[K]Q[DKL]R[EKD]E[KR]V[A]T[AEQ]R[EKD]E[L]E[KNS]E[KR]I[L]AK[E]NS[A]D[KE]T (SEQ ID NO: 337)	[T]LE[KNR]K[EQR]I[L]AK[RQ]NS[A]D[E]T (SEQ ID NO: 338)	R[KED]V[A]R[KQE]E[KQ]L[QK]E[K]E[KDQ]R[KDE]I[L]R[K]K[D]N[H]SD[E]K (SEQ ID NO: 339)
DHR69 _design	NPQEDLERAEKVVRVVEEV LQRAKEAQREGDKEKVERL IKEAENQIRKARELLERVVR QNPDD (SEQ ID NO: 346)	PEVLLRVAELIVRLVEVVLE LAKLAEKNGDKEQVERLIQ TAEELIREARELLERSREIP DN (SEQ ID NO: 347)	PESLKRVAELIKRLVKVVDE LSKLAERNGDDRQVERLRQ LAEELRREAELEERVRRER PD (SEQ ID NO: 348)
DHR69 _variant s	N[D]P[S]Q[EDK]E[DK]D[ELK] I[L]A[E]K[R]R[K]E[A]E[KR]K[E] Q[V]L[V]A[R]K[E]S[K]E[V]A[I] E[KHQ]E[KR]V[AD]I[L]AIV[Q] [ER]K[R]KDE[AK]E[R]E[KQR] AQ[TS]R[K]E[KD]GD[N]K[E] E[E]D[T]K[ET]R[V]A[E]R[KQ]R [KE]L[R]I[T]K[E]E[KR]AE[D] RS[N]E[KQ]Q[LKA]I[V]R[KE] Q[K]RE[AR]KTE[E]KT[L]AE K[LE]RDQ[R]K[E]V[A]V[AKR] [R]K[N]Q[KDN]N[RA]D[PD]T D[N] (SEQ ID NO: 343)	P[W]A[E]K[DQ]V[AL]LL[A]R[EKQ]V[I]AE[KRQ]L[I]V[A] R[EDK]LV[A]E[R]K[N]V[A]IN] V[A]I[L]AIV[E]R[K]LAK[E]E [AE]Q[A]K[N]EQ[N]EDT]GD[N] [K]E[E]DK]Q[KET]V[A]E[RH] Q[R]EKQ]L[I]D]Q[EKR]T[EQ D]AE[KQS]E[R]K[L]DAI]I[V] R[KEQ]E[KD]AR[E]K]E[KR] L[A]LE[DR]K[R]KQE]V[A]S[A] KR]R[KND]E[N]DQ]I[RA]D]PD [T]N[D] (SEQ ID NO: 344)	P[W]A[E]K[DKH]S[AL]LK[QR]R [KE]V[I]AE[KQ]L[I]L]K[ED R]R[EK]LV[A]R[K]E[V]A]IN] V[A]I]D[EK]E[KR]L[Q]S[A]K E]L[S]AE[KQ]R[K]N]EST]GD [N]R]EST]D[E]Q[KTD]V[A]E KRN]R[KET]LR]KEN]Q[KE]L [EQ]T]AE[KR]E[KRQ]L[DAI]R [EK]R[K]E[KDQ]AE[KR]E KQ]L[A]E[KQ]E[KR]R[ILD]V [A]R[K]E[R]KND]E[N]QT]R[A DQ]PD[T] (SEQ ID NO: 345)
DHR70 _design	STEEKIEEARQSIKEAERSLR EGNPEKAREDVRRALELVR ELEKLARKTGS (SEQ ID NO: 352)	TEVLIEAARLAIEVARVALK VGSPETAREAVRTALELVQE LERQARKTGS (SEQ ID NO: 353)	DEVLKRAELAKEVARVAK EVGSPETARQARETAERLRE ELRRNREKKG (SEQ ID NO: 354)
3DHR7 0_variant ts	S[DN]T[SDN]E[DK]E[DO]K[L RT]I[ALW]E[KQ]E[DKS]AR EKQ]Q[KRD]S[A]I[V]K[ER]E [KQR]AE[QK]R[K]E[S]ADN]L [AHR]R[KED]E[KQR]GN[SD] P[DKS]E[DKQ]K[STE]AR[EK [E]K[RQ]D[A]V[AL]R[K]EQ[R [KE]AL[E]Q]E[KDN]L[A]V[A R]R[K]EQ]E[KR]L[A]E[S]AI[K] ERQ]L[ERD]AR[KEQ]K[ERT] T[QRK]G[D]S[DN] (SEQ ID NO: 349)	T[I]DV[E]RDK]V[A]L[I]ALV[E [KAD]AAR[EK]L[I]A]V[E]R KQ]V[A]AR[EKD]V[A]AL[A Q]K[ERN]V[T]GS]D]P[ST]E DQ]T[L]V[AR]E[E]K]AV[AL] R[EKQ]T[QLE]ALE[KNQ]L[A I]V[A]Q[R]KE[E]RD[L]A]E[A KR]R[K]E]Q[EAR]AR[EK]K[R E]T[SEH]GS[DN] (SEQ ID NO: 350)	D[T]IR]E[DKQ]V[AT]LK]EQR [R]EKT]AAE[RKD]L[I]AK[ER [E]KQR]V[A]AR[EK]V[A]AK QER]E[KQR]V[T]GS]D]P[SD] E]D]T[LSV]AR[KEQ]Q[KED] AR[EKQ]E[KQR]T[LQ]AE KRQ]R[KDN]L[A]I[R]EKD[E] KQ]E[RQ]A]L[IA]R[K]E[R]K ED]N]EQ]A]R]ADN]E[KR]K[R [K]QRN]G (SEQ ID NO: 351)
DHR71 _design	DPEEILERAKESLERAREASE RGDEEFFRKA AEKALELAK RLVEQAKKEGD (SEQ ID NO: 358)	PELVLEAAKVALRVAELAA KNGDKEVFKKAAESALEVA KRLVEVASKEGD (SEQ ID NO: 359)	PELVLEAAKVAEEVRKLAK KQGDEEVYEKARETAREVK EELKRVREEKG (SEQ ID NO: 360)
DHR71 _variant s	D[N]P[SD]E[D]E[DR]I[TVD]L [AET]E[K]R[K]N]AK[RE]Q]E KR]S[AE]L[E]RDK]R[KET]AR [E]E[KQ]AS[AHK]E[K]N]R[K DQ]GDE[DSQ]E[DKQ]E[TK]F R[KQ]K[EDR]AAE[RKQ]K[R TN]ALE[KDR]L[ITV]AK[QRE R]K[L]A]V[A]E[KD]Q[ERK] AK[ERS]K[ENQ]E[QDK]GD N] (SEQ ID NO: 355)	P[A]E[KQ]L[A]V[I]L]A]E[DK] AAK[RE]Q]V[T]ALR[EK]V[L] AE[R]LAA[K]K[ER]N[KQE]G DK[DSQ]E[DKQ]VFK[QR]K[E DQ]AAE[KRD]S[TAV]ALE[K T]V[I]L]AK]QE]R[ED]L[A]V A]E[KR]V[EQ]AS[KER]K[NE] E[Q]GD[N] (SEQ ID NO: 356)	P[ALT]E[DN]L[A]V[I]E[ERKQ] E[QKL]AAK[ERQ]V[T]AE[KR [E]R[K]V[L]R[A]K]ER[LAK]E R[K]E]Q[KRE]GDE[DRS]E[D] V[L]Y[FR]E[K]K[ERQ]AR[EQ [E]KDR]T[V]A]AR[E]E[KRT]V [IL]K[ETR]E[K]E[R]L[A]K]E [R]EKH]V[EQ]R[A]E]K]T]E[K NR]K[QE]G (SEQ ID NO: 357)
DHR72 _design	DSTKEKARQLAEAEAKETA KVGDPelikLAEQASQEGD (SEQ ID NO: 364)	SEKAKAILLAEEAARVAKE VGDPelikLAEEAARRGD (SEQ ID NO: 365)	SEKARAILEAERAREAKER GDPEQIKKARELAKRG (SEQ ID NO: 366)
DHR72 _variant s	D[N]S[TD]T[DE]K[ETS]E[DK Q]K[ERD]AR[K]Q[EDK]L[RK [AE]KND]E[K]AK[AQ]E[KH] T]I]VS]AE[K]K[ER]V[TA]GD NS]PE[DH]N]L[K]K[ER]L[T] AE[KQD]Q[EKR]AS[A]Q[KD	S[R]E[KD]K[W]AK[ER]AI]V A]L[K]L[R]AAE[K]AAR[KEL V]I]T]AKE[KQ]V[T]GD]NS]P E]D]L[K]R]LAL[REQ]E[KQ]A AR[KE]R[EDN]GD (SEQ ID NO: 362)	S[AKR]E[DKR]K[QW]AR[ED [AI]VA]L[KR]E[R]K]AAE[KR] R[KET]AR[KLE]E[K]AKE[KQ [R]EK]GD[SN]PIS]E[DNQ]Q KRT]K[EQ]K[ER]AR[EKQ]E KR]L[EK]AK[RE]Q]R[EK]G

	R E DQR GD N (SEQ ID NO: 361)		(SEQ ID NO: 363)
DHR73 _design	DAEEEEAKEAIKRAQEAIELA RKGNPPEARKVAEEARERA ERVREEAEKRGD (SEQ ID NO: 370)	AEVLALVAIALALVAIALAE VGNPEEAREVAERAKEIAER VRELAEKRGD (SEQ ID NO: 371)	ARVLKLVAKALELVAEALK KVGNPPEAREVEERAREIKE RVRRLLEEKG (SEQ ID NO: 372)
DHR73 _variant s	D N S A SD E R E D E K R A K E K A K E R D K A Q K E R K A I S E K L K D E A R K Q E K R G N D P E D K E S R T A R K E K E V T K E A E D R E D Q R A R E Y E K R R I L D A E Q D K R E K V A R E A L E K R E R K N A E R K Q K E R R K Q G D N S (SEQ ID NO: 367)	AE RK V A L L V A I A L A L V A I A L A E Q K V G N D P E D E S A R E Y K E R K V A E R D R E D T A K R Y E E K R Q I L V A E Q D R R E V A R E Y E K R L E Q A E R Q K E R R Q D N G D N (SEQ ID NO: 368)	A D I W R D E K V A L K E Q R L V A K E R A L E K L V A E K A L K R Q K Q E N V G N D P E D E S A R E K T E K R S V E K Q E R K Q R Q D E A R E Q K E K R I V L T K Q E R E K R D R E D K V A R K D E R K E Q L E I N L A K E K R T E K R K R Q N G (SEQ ID NO: 369)
DHR74 _design	DSEADRIKKLQKEIKEVEQE ARDSNDDEERELLKRLAEA LKRAAEAVKRAQESGD (SEQ ID NO: 376)	SEAIRIHKLLVKEITEVVREA RKSTDKEIEILLRLAEALAR AAEAVADAAGSGD (SEQ ID NO: 377)	QEAIRIKKLLVKKIIEVVVRK ARKSTNKKEIEKLIRKAEKL ARKAEQIAEDAKRG (SEQ ID NO: 378)
DHR74 _variant s	D N S T D N E D Q T A D K E N R K E I L I A R K E D K R Q L Q K E K R E E A Q I K E D E K R V I L E Q K Q E K R E K N Q A R K E N D K R E S E A R N T D N D S P Q E T D E K L Q R I Q E K D L L K Q R K L L A E A L A K Q E R R I A A E D K R A V A K Q D E R I E K A Q A R E E K D Q S T Q A G D (SEQ ID NO: 373)	S Q D E E Q R A I L R K E D I L I K R K E Q S L V A K E H R E A D L I T I L E K R V I L V A I R E Q K E R A R D E T K R S A E Q I D K E P Q E D N E R K I E K H R L L I V R K D L L A E A A A R A A E A V A A D K R E A A K E Q S T A K G D N (SEQ ID NO: 374)	Q E D E D S A I L K E D R K T I K E K E Q R L V A K E R K R N I L S E K Q V I L K V A I R E K Q K E Q R A R E K N K I R E N S A E K I T N N D K E P Q K E T D E K Q R E K R K E L K R I V R E K Q K E A E K Q K E D L A A R D K K E A E Q E N R I A E L A E K R D R K A K E Q R R K E D G (SEQ ID NO: 375)
DHR75 _design	DSEKEKATELAERAQDVAS RVEEEARREGSRELIEIAREL RERAEASQEGD (SEQ ID NO: 382)	SEKAKAILLAAKAVLVAVE VYERAKRQGSDELREIAREL AKEALRAAQEGD (SEQ ID NO: 383)	SEKARAILEAAREVLRAVEQ YERAKRRGDDDERERAREE AREALERAREG (SEQ ID NO: 384)
DHR75 _variant s	D N S D N T E D K T K E S E D K K E E R T A T K R E K R H L K E R A E K N R K E A Q I K D K E V L T I A S K E Q R E K Q V A E L K R E K R E L R A R D K Q R K E Q E T D Q G S R D S E D K L I E K A E K Q N I A R E K Q E K Q R L R A E E K R R L E Q A E K E K Q R A S A Q K R E E R K G D N (SEQ ID NO: 379)	S A P E E K D S K A R A K E L D A I V L A K R L K R E A A K E D L A V L K R A V I L T A V I A E Q R V A Y E R K R L E K A K R H R E K Q Q E N G S D E S E D T K L R K Q E K N Q I A R E K Q E R K Q L A K R E E L A R K E Q A A Q K R E R G D (SEQ ID NO: 380)	S A P E E D K S K R A R E D Q A I V L A K R E K Q A A R K E Q E K R A R V L K R A R E K Q A V I A E Q R K Q E K N Y E S K A R K E T A K R D S R K E R K E G D S D E S D E E K D R Q A E E R K Q R K E A R E K N E R D K E K R A R K E E K Q A L E K R R E K R A R K Q E R G (SEQ ID NO: 381)
DHR76 _design	NPELEEWIRRAKEVAKEVE KVAQRAEEEGNPDLRDSAK ELRRAVEEAIEEAKKQGN (SEQ ID NO: 388)	PELVWVARAAKVAEEVIK VAIQAEKEGNRDLFRAALEL VRVIAEIEEAVKQGN (SEQ ID NO: 389)	PELVERVARLAKKAAELIKR AIRAEKEGNRDERREALERV REVIERIEELVRQG (SEQ ID NO: 390)
DHR76 _variant s	N D S P N D S E K D R L RT E K Q E K W A I D K R K D R E K A K Q E N E K R Q V A A K E N D E K D V A E K Q R K E V L Q A A Q K E R K A I L E K Q R E K R N E N S Q G N D P D E D E K L R A T D R N E S A L I A K E N R E K R L R V K R E K D A V I E Q R K E R A E K R E Q R A K Q R S K R E N Q E R G N D S (SEQ ID NO: 385)	P W A S E K R D L V A E K R W A V A R E K A A K E Q R V A A A V E K L R V A I L K E Q R V L Q A I E L Q K R A L E Q K A K N E D S N G N R P E K D K E T L F A R T R K E D A L V A L A R E K R L V R E K A V I E R K A E K R E K R A V A K D E Q K G N D S (SEQ ID NO: 386)	P W A E K D L V A E K Q R R E K T V A R E K D L R E K A K E R K E Q A A V E K Q L V A E I L K E Q R K E H A I L E R E K A D L E Q K H K N R Q E N R K G N D R P E K D E K E K R D R A T R E K D E K R A N L A E K E K R R K E T V R E K D E K N V I E K R Q R T E K I E K E K L A S V A R K D S Q E K R G (SEQ ID NO: 387)

DHR77 _design	NSDEEEAREWAERAEAAK EALQAKREGDEDARRVAE ELEKQAEARRKKD (SEQ ID NO: 394)	SEEAEAVYWAARAVLAAL ALEQAKREGDEDARRVAE LLRQAEAAARKKN (SEQ ID NO: 395)	PEEARAVYEAARDVLEALQ RLEEAKRRGDDEERREAEER LRQAEERARKK (SEQ ID NO: 396)
DHR77 _variant s	N[D]S[T]D[ER]E[DK]E[KDQ] E[K]N[AR]QK[E]KQR[W]A[A] V[E]DRK[R]EK[AE]KR[E]RK [A]L[AK]Q[RD]E[KR]AL[EK] E[K]Q[EKL]AK[QRE]R[K]E QR[GDE]D[QRE]AR[LE]R [KE]V[L]AE[KQD]E[RQ]LE[R] L[K]ER[Q]ELR[AE]KRD[E]K AR[EKA]R[EK]K[N]K[NH]Q D[NS] (SEQ ID NO: 391)	S[ARQ]E[DKS]E[LT]AE[AKQ] [AVY]A[W]A[A]V[AR]LEK[A] V[A]L[A]A[L]ALE[KOR]ALE [L]Q[L]AK[Q]R[E]E[Q]GDE D[D]QK[AR]IQE[R]EK[V]L[A] E[RKQ]E[RK]LLR[KE]Q[L]A E[R]E[K]AA[L]R[E]K[N]K N[N]D] (SEQ ID NO: 392)	P[KAE]E[KDS]E[TQD]AR[ED] N[AVY]A[E]KRD[A]V[AR]K E[D]AEK[V]A[L]YAK[E]KR ALQ[ERK]R[E]K[L]Y[E]HIK[E] [KOR]AK[ER]R[K]R[KED]GD [N]E[DKQ]E[DK]E[ADK]R[K] Q[R]KEQ[E]KRS[AE]K[R]E[K DR]R[EKN]LR[KE]Q[KER]AE [KR]E[KR]R[AKN]A[Q]R[ED] K[K]NR[K]NEH] (SEQ ID NO: 393)
DHR79 _design	SSDEEEARELIERAKEAAER AQEAAERTGDPVRVRELARE LKRLAQEAAEEVKRDPSS (SEQ ID NO: 400)	SDVNEALKLIVEAIEAAVRA LEAAERTGDPVRELARELV RLAVEAAEEVQRNPSS (SEQ ID NO: 401)	EEVNEALKKIVKAIQEAVES LREAEESGDPKREKARERV REAVRAEEVQRDPS (SEQ ID NO: 402)
DHR79 _variant s	S[ND]S[DTN]D[E]E[DK]E[KD] [E]KRT[AR]EK[E]KR[L]RAE [TK]E[RK]R[KE]AK[EQ]E[K R]AA[S]E[KRD]R[EKL]AQ[E KN]E[QR]AAE[KNR]R[EKN] T[A]GDPR[KNT]V[A]R[IK]E K[LAR]KE[E]K[R]LK[SRV]R ED]LAQ[EK]R[E]RKN]AAE[K R]E[QRD]VK[Q]E[R]K]DPS[R T]S[ND] (SEQ ID NO: 397)	S[DKE]D[N]V[A]N[RV]E[RK] AL[A]K[ED]L[R]V[V]IL[E]K R[A]E[K]AAVR[EAK]ALE[K] AAE[IKN]R[KQ]T[V]GDPE[K RN]V[A]R[IE]K[LAR]AV[E] KR]LVR[EQ]LAVE[RK]AAE K[E]RKN]VQ[WLD]R[E]K[N] D]PS[RK]S[DN] (SEQ ID NO: 398)	E[DKQ]E[DS]V[AS]N[V]A[E]R DK]AL[A]K[E]K[ER]V[V]IL [K]RQ[A]L[Q]EKD[E]DK[AV E[RKQ]S[A]LR[E]KQ[E]KNR] AE[NKQ]E[RDK]S[KTE]GD N]PE[NQ]K[EQ]R[IKQ]E[K]K [RE]AR[AV]E[KR]R[E]KQ]VR [E]E[KR]AVE[RK]R[KET]AE QK[E]K[V]I[Q]HLA]R[KN]DP S[NRT] (SEQ ID NO: 399)
DHR80 _design	NSEELERESEEAERRLQEAR KRSEEARERGD[L]KELAEALI EEARAVQELARVASERGN (SEQ ID NO: 406)	SEEAEERASEKAQRVLEEAR KVSEEARERQGDDEV[L]ALALI AIALAVLALAEVASSRGN (SEQ ID NO: 407)	KEEAERAYEDARRVEEAR KVKESEAEQGDSEVKRLAE EAEQLAREARRHVQETRG (SEQ ID NO: 408)
DHR80 _variant s	N[SD]S[TD]E[DK]E[DKQ]L[A] D[E]KQ[R]KE[E]RN[S]EAH E[KR]E[KD]AE[KQR]R[KE]R [EKD]L[YAE]Q[REK]E[KR]A R[KE]K[E]R[E]K[S]A[E]K[E]K R[AR]KEQ[E]K[R]R[KQT]GD LK[RE]Q[E]TAK[L]AEK[AE]K [ALIE]KRI[E]RIA]AR[QKD]A V[A]Q[KA]E[E]KIR]L[IAK]AR [EK]V[A]AS[KAE]E[RKD]R KEA]GN[SD] (SEQ ID NO: 403)	S[RDK]E[DKQ]E[TLA]AE[DK] Q[R]EKQ[AS]AEK[E]K[R]K[R EW]AQ[E]KR[R]KQ[E]VL[YAE E]K[RQ]E[QDK]AR[E]KQ]K[E V]I[S]A[E]KRD]E[KRQ]AR[E K]E[K]Q[KEN]GD[N]D[LYE] E[RQK]V[A]L[A]ALALIA]A RQ]AL[Q]AV[A]L[AV]AL[IK A]AE[ILV]V[A]AS[AEK]S[A] R[EKS]GN[DS] (SEQ ID NO: 404)	K[QSR]E[DKS]E[KTA]AE[KD R]R[E]K[AY]AEK[E]K[RQ]D[K ER]AR[E]KQ]R[E]KQ]VE[KYA E]K[R]E[RKS]AR[E]KS]K[E]V [IK]AR[E]R[K]S[EQ]R[AE]KD Q]E[KR]Q[KN]GD[N]S[DEQ] E[K]V[A]K[L]YA]R[KND]LAE [KNQ]E[TKR]AE[IA]R[Q]EKN [L]K[AR]EKD]E[RK]AR[KA] R[KQ]E[H]ILQ]V[A]Q[KAD]E KRD]T[SA]R[KQS]G (SEQ ID NO: 405)
DHR82 _design	NDEEVQEAVERAEELREEA EELIKKARKTGDPPELLRKAL EALAEAVRAVEEAIKRNPON (SEQ ID NO: 412)	DEAVETAVRLARELKKVAE ELQERAKKTGDPPELLKLAL RALEVAVRAVELAIKSNPD N (SEQ ID NO: 413)	EEAVETAKRLAEELRKVAE LLEERAKETGDPPELLQELAKR AKEVADRARELAKKSNPN (SEQ ID NO: 414)
DHR82 _variant s	N[D]D[TS]E[DR]E[DR]V[AD] Q[KE]E[KRQ]A[KL]V[ANQ]E [RK]R[KDE]AE[QRD]E[RQK] L[A]R[AE]E[KR]E[KD]AE[R KQ]E[K]L[AW]I[AEQ]K[DER K]E[AR]KEQ]K[E]T[E]K[G[N] D]TNS]P[T]E[DTQ]L[AK]LR KDE[K]E[AL]IVAE]R[A]KR W]L[V]E[IKR]E[KR]AV[A]R EK]A[L]V[A]E[AKR]E[QKR] AI[L]K[R]D[R]DOE]N[DER]P	D[EKS]E[DK]AV[A]E[KNR]T [AIL]A[L]V[ANQ]R[EDK]L[I AV]AR[E]KQ]E[LDR]L[A]K[A Q]K[ER]V[A]I[AE]KQ]E[DK L]L[A]Q[IA]E[KQ]R[R]LE]A K[RE]Q[K]RD]T[E]GD[NTS]P T]E[QRT]L[A]LK[RDE]L[E]K W]AL[IVAE]R[E]K[AL]V[E]K V[AL]AV[A]I[R]KEQ]AV[A]E A[L]AE]AI[L]K[RE]S[A]N[D RE]PD]N[SE]N[D] (SEQ ID	E[KQ]R[E]DRK]AV[A]E[KDR] T[AIL]A[L]K[QEA]R[E]KS]L[I AV]AE[KQ]R[E]LRA]L[A]R[K EQ]K[RES]V[A]I[AE]K[R]L[DE I]L[A]E[K]E[KRD]R[LIQ]AK QER]E[KR]T[HQ]G[N]D[TNS] P[ERS]E[TDR]L[A]Q[E]E[K N]L[AK]I[AK]EQ[R]EKD]AK ER]E[KR]V[AL]AD[KNR]R[E K]AR[E]KQ]E[KQ]L[AE]I[AK] RQ]K[RDE]S[A]N[DRS]P[N]D

	D[RGN]N[D] (SEQ ID NO: 409)	NO: 410)	ST] (SEQ ID NO: 411)
--	-----------------------------	----------	----------------------

In another embodiment, the polypeptide comprises or consists of the amino acid sequence selected from the group consisting of:

- (A) SEQ ID NO:4-[SEQ ID NO:5]_(0 or 2-19)-SEQ ID NO:6;
- 5 (B) SEQ ID NO:10-[SEQ ID NO:11]_(0 or 2-19)-SEQ ID NO:12;
- (C) SEQ ID NO:16-[SEQ ID NO:17]_(0 or 2-19)-SEQ ID NO:18;
- (D) SEQ ID NO:22-[SEQ ID NO:23]_(0 or 2-19)-SEQ ID NO:24;
- (E) SEQ ID NO:28-[SEQ ID NO:29]_(0 or 2-19)-SEQ ID NO:30;
- (F) SEQ ID NO:34-[SEQ ID NO:35]_(0 or 2-19)-SEQ ID NO:36;
- 10 (G) SEQ ID NO:40-[SEQ ID NO:41]_(0 or 2-19)-SEQ ID NO:42;
- (H) SEQ ID NO:46-[SEQ ID NO:47]_(0 or 2-19)-SEQ ID NO:48;
- (I) SEQ ID NO:52-[SEQ ID NO:53]_(0 or 2-19)-SEQ ID NO:54;
- (J) SEQ ID NO:58-[SEQ ID NO:59]_(0 or 2-19)-SEQ ID NO:60;
- (K) SEQ ID NO:64-[SEQ ID NO:65]_(0 or 2-19)-SEQ ID NO:66;
- 15 (L) SEQ ID NO:70-[SEQ ID NO:71]_(0 or 2-19)-SEQ ID NO:72;
- (M) SEQ ID NO:76-[SEQ ID NO:77]_(0 or 2-19)-SEQ ID NO:78;
- (N) SEQ ID NO:82-[SEQ ID NO:83]_(0 or 2-19)-SEQ ID NO:84;
- (O) SEQ ID NO:88-[SEQ ID NO:89]_(0 or 2-19)-SEQ ID NO:90;
- (P) SEQ ID NO:94-[SEQ ID NO:95]_(0 or 2-19)-SEQ ID NO:96;
- 20 (Q) SEQ ID NO:100-[SEQ ID NO:101]_(0 or 2-19)-SEQ ID NO:102;
- (R) SEQ ID NO:106-[SEQ ID NO:107]_(0 or 2-19)-SEQ ID NO:108;
- (S) SEQ ID NO:112-[SEQ ID NO:113]_(0 or 2-19)-SEQ ID NO:114;
- (T) SEQ ID NO:118-[SEQ ID NO:119]_(0 or 2-19)-SEQ ID NO:120;
- (U) SEQ ID NO:124-[SEQ ID NO:125]_(0 or 2-19)-SEQ ID NO:126;
- 25 (V) SEQ ID NO:130-[SEQ ID NO:131]_(0 or 2-19)-SEQ ID NO:132;
- (W) SEQ ID NO:136-[SEQ ID NO:137]_(0 or 2-19)-SEQ ID NO:138;
- (X) SEQ ID NO:142-[SEQ ID NO:143]_(0 or 2-19)-SEQ ID NO:144;
- (Y) SEQ ID NO:148-[SEQ ID NO:149]_(0 or 2-19)-SEQ ID NO:150;
- (Z) SEQ ID NO:154-[SEQ ID NO:155]_(0 or 2-19)-SEQ ID NO:156;
- 30 (AA) SEQ ID NO:160-[SEQ ID NO:161]_(0 or 2-19)-SEQ ID NO:162;
- (BB) SEQ ID NO:166-[SEQ ID NO:167]_(0 or 2-19)-SEQ ID NO:168;
- (CC) SEQ ID NO:172-[SEQ ID NO:173]_(0 or 2-19)-SEQ ID NO:174;

(DD) SEQ ID NO:178-[SEQ ID NO:179]_(0 or 2-19)-SEQ ID NO:180;
 (EE) SEQ ID NO:184-[SEQ ID NO:185]_(0 or 2-19)-SEQ ID NO:186;
 (FF) SEQ ID NO:190-[SEQ ID NO:191]_(0 or 2-19)-SEQ ID NO:192;
 (GG) SEQ ID NO:196-[SEQ ID NO:197]_(0 or 2-19)-SEQ ID NO:198;
 5 (HH) SEQ ID NO:202-[SEQ ID NO:203]_(0 or 2-19)-SEQ ID NO:204;
 (II) SEQ ID NO:208-[SEQ ID NO:209]_(0 or 2-19)-SEQ ID NO:210;
 (JJ) SEQ ID NO:214-[SEQ ID NO:215]_(0 or 2-19)-SEQ ID NO:216;
 (KK) SEQ ID NO:220-[SEQ ID NO:221]_(0 or 2-19)-SEQ ID NO:222;
 (LL) SEQ ID NO:226-[SEQ ID NO:227]_(0 or 2-19)-SEQ ID NO:228;
 10 (MM) SEQ ID NO:232-[SEQ ID NO:233]_(0 or 2-19)-SEQ ID NO:234;
 (NN) SEQ ID NO:238-[SEQ ID NO:239]_(0 or 2-19)-SEQ ID NO:240;
 (OO) SEQ ID NO:244-[SEQ ID NO:245]_(0 or 2-19)-SEQ ID NO:246;
 (PP) SEQ ID NO:250-[SEQ ID NO:251]_(0 or 2-19)-SEQ ID NO:252;
 (QQ) SEQ ID NO:256-[SEQ ID NO:257]_(0 or 2-19)-SEQ ID NO:258;
 15 (RR) SEQ ID NO:262-[SEQ ID NO:263]_(0 or 2-19)-SEQ ID NO:264;
 (SS) SEQ ID NO:268-[SEQ ID NO:269]_(0 or 2-19)-SEQ ID NO:270;
 (TT) SEQ ID NO:274-[SEQ ID NO:275]_(0 or 2-19)-SEQ ID NO:276;
 (UU) SEQ ID NO:280-[SEQ ID NO:281]_(0 or 2-19)-SEQ ID NO:282;
 (VV) SEQ ID NO:286-[SEQ ID NO:287]_(0 or 2-19)-SEQ ID NO:288;
 20 (WW) SEQ ID NO:292-[SEQ ID NO:293]_(0 or 2-19)-SEQ ID NO:294;
 (XX) SEQ ID NO:298-[SEQ ID NO:299]_(0 or 2-19)-SEQ ID NO:300;
 (YY) SEQ ID NO:304-[SEQ ID NO:305]_(0 or 2-19)-SEQ ID NO:306;
 (ZZ) SEQ ID NO:310-[SEQ ID NO:311]_(0 or 2-19)-SEQ ID NO:312;
 (AAA) SEQ ID NO:316-[SEQ ID NO:317]_(0 or 2-19)-SEQ ID NO:318;
 25 (BBB) SEQ ID NO:322-[SEQ ID NO:323]_(0 or 2-19)-SEQ ID NO:324;
 (CCC) SEQ ID NO:328-[SEQ ID NO:329]_(0 or 2-19)-SEQ ID NO:330;
 (DDD) SEQ ID NO:334-[SEQ ID NO:335]_(0 or 2-19)-SEQ ID NO:336;
 (EEE) SEQ ID NO:340-[SEQ ID NO:341]_(0 or 2-19)-SEQ ID NO:342;
 (FFF) SEQ ID NO:346-[SEQ ID NO:347]_(0 or 2-19)-SEQ ID NO:348;
 30 (GGG) SEQ ID NO:352-[SEQ ID NO:353]_(0 or 2-19)-SEQ ID NO:354;
 (HHH) SEQ ID NO:358-[SEQ ID NO:359]_(0 or 2-19)-SEQ ID NO:360;
 (III) SEQ ID NO:364-[SEQ ID NO:365]_(0 or 2-19)-SEQ ID NO:366;
 (JJJ) SEQ ID NO:370-[SEQ ID NO:371]_(0 or 2-19)-SEQ ID NO:372;
 (KKK) SEQ ID NO:376-[SEQ ID NO:377]_(0 or 2-19)-SEQ ID NO:378;

(LLL) SEQ ID NO:382-[SEQ ID NO:383]_(0 or 2-19)-SEQ ID NO:384;
 (MMM) SEQ ID NO:388-[SEQ ID NO:389]_(0 or 2-19)-SEQ ID NO:390;
 (NNN) SEQ ID NO:394-[SEQ ID NO:395]_(0 or 2-19)-SEQ ID NO:396;
 (OOO) SEQ ID NO:400-[SEQ ID NO:401]_(0 or 2-19)-SEQ ID NO:402;
 5 (PPP) SEQ ID NO:406-[SEQ ID NO:407]_(0 or 2-19)-SEQ ID NO:408; and
 (QQQ) SEQ ID NO:412-[SEQ ID NO:413]_(0 or 2-19)-SEQ ID NO:414;
 wherein the domain in brackets is an optional internal domain.

The polypeptides of this embodiment include 2 or 3 domains (as described above),
 and are represented in Table 1 above, reflected in each row showing listed as
 10 "DHRx_design" (where x is replaced by a specific number in the table).

In one embodiment of any aspect or embodiment of the polypeptides, the internal
 domain is absent. In certain alternative embodiments, the polypeptides according to this
 aspect further comprise at least one of an N_{cap} domain coupled to the N-terminus of the at
 least two Internal domains and a C_{cap} domain coupled to the C-terminus of the at least two
 15 Internal domains. In certain embodiments, the optional internal domain is present in 2-19
 copies. In certain specific embodiments, the optional internal domain is present in 2-3
 copies.

In another aspect, the invention provides polypeptides comprising or consisting of a
 polypeptide having at least 50% identity over its length with a polypeptide having the amino
 20 acid sequence selected from the group consisting of SEQ ID NO: 415-497 (see Table 2). The
 polypeptides of this aspect of the invention represent novel repeat proteins with precisely
 specified geometries identified using the methods of the invention, opening up a wide array
 of new possibilities for biomolecular engineering. In various embodiments, the polypeptides
 comprise or consist of a polypeptide having at least 55%, 60%, 65%, 70%, 75%, 80%, 85%,
 25 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity over its length with a polypeptide having
 the amino acid sequence selected from the group consisting of SEQ ID NO: 415-497.

TABLE 2.

Name	Sequence
DHR1	GCDQVAKDASSTIREVIEKNPNYSEKVADVAAKIVKKIIEGNPNGC DCVAKAASSIIRAVIEKNPNYSEVVADVAAAIKAIIEGNPNGCDCVA KAASSIIRAVIEKNPNYSEVVADVAAAIKAIIEGNPNGRDCVRKAAS SIIRAVQEKNPNYSEVVEDVKRAIEKAIKEGNPN (SEQ ID NO: 415)
DHR2	SDADEAAKEANKAENKARNRNDDEAAKAVKLIKEAIERAKKRNESD

	AVEAAKEAAKALNKKALNRNDDEAAKAVALLIAEAIIRALKRNESDAVE AAKEAAKALNKKALNRNDDEAAKAVALLIAEAIIRALKRNESDAVEKAK EAAKNLKKALNRNDDEQAKHVAKQAENIRALKRNES (SEQ ID NO: 416)
DHR3	SSEDTVRLIAQKCSEAIRESNDCEEAAARKCAKTISEAIRESNSSSELAVRI IAQVCSEAIRESNDCECAARICAKIISEAIRESNSSSELAVRIIAQVCSEAIR ESNDCECAARICAKIISEAIRESNSSSELAKRIKQVCSEAKRESNDTECA KRICKIKSEAKRESNS (SEQ ID NO: 417)
DHR4	SYEDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESG SSYEVICCVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGS SYEVICCVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSS YEVIKECVQRIVEEIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSS (SEQ ID NO: 418)
DHR5	SSEKEELRERLVKICVENAKRKGDDTEEAAREAAEFELVREAAERA GIDSSEVLELAIRLIKECVENAQREGYDISEACRAAAAEAFKRVAEAAK RAGITSSEVLELAIRLIKECVENAQREGYDISEACRAAAAEAFKRVAEAA KRAGITSSETLKRAIEEIRKRVEEAQREGNDISEACRQAAEEFRKKAEE LKRRGD (SEQ ID NO: 419)
DHR6	SEEKEEALKKVREAAKKLGSSDEEARKCFEEAREWAERTGSSAYEAA EALFKVLEAAAYKLGSSAEACECFNQAAEWAERTGSGAYEAAEALFK VLEAAAYKLGSSAEACECFNQAAEWAERTGSGAYEAAERLFEELERA YEEGSSAEACEEFNKKEEEAHRKGKK (SEQ ID NO: 420)
DHR7	STKEDARSTCEKAARKAAESNDEEVAKQAAKDCLEVAKQAGMPTKE AARSFCEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKEAARS FCEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEA AKRAAKESNDEEVEKIAKKACKEVAKQAGMP (SEQ ID NO: 421)
DHR8	SDEMKKVMEALKKAVELAKKNNDDEVAREIARAAAEIVEALRENNS DEMAKVMLALAKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSD EMAKVMLALAKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSDE MAKKMLELAKRVLDAAKNNDDETAREIARQAAEEVEADRENNS (SEQ ID NO: 422)
DHR9	SYEDEAEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESG SSYEVIAEIVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSS YEVIAEIVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSY EVIKEIVQRIVEEIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSS (SEQ ID NO: 423)
DHR10	SSEKEELRERLVKIVVENAKRKGDDTEEAAREAAEFELVREAAERA GIDSSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAK RAGITSSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEA AKRAGITSSETLKRAIEEIRKRVEEAQREGNDISEAARQAAEEFRKKAEE

	ELKRRGD (SEQ ID NO: 424)
DHR11	SDADEAAKEANKAENKARNRNDDEAAKAVKLCKEAIERAKKRNESD AVEAAKEAAKALNKALNRNDDEAAKAVALCCEAIIRALKRNESDAV EAAKEAAKALNKALNRNDDEAAKAVALCCEAIIRALKRNESDAVEK AKEAAKNLNLKALNRNDDEQAKHVAKQCENIIRALKRNES (SEQ ID NO: 425)
DHR12	DDEEQCREIAEKAKQTYTDDDEIARIHAEAAARQTTTDDDEICRCIAEAA KQTYTDDDEIARIHAYAAARQTTTDDDEICRCIAEAAKQTYTDDDEIARIH AYAAARQTTTDDDEIERCIEEAAKQTYTDDDEIERIKEYARRQTTTDD (SEQ ID NO: 426)
DHR13	NAEDKAREVLKELKDEGSPEEEAARQVLKDLNREGSNAEDAARAVL KALKDEGSPEEEAARAVLKALNREGSNAEDAARAVLKALKDEGSPEE EAARAVLKALNREGSNEEDASRAVLKALKDEGSPEEEARRAVEKALN REGSN (SEQ ID NO: 427)
DHR14	DSEEVNERVKQLAEKAKEATDKKEEVIEIVKELAEAKQSTDSELVNEI VKQLAEVAKEATDKELVIYIVKILAEAKQSTDSELVNEIVKQLAEVA KEATDKELVIYIVKILAEAKQSTDSELVNEIVKQLEEVAKKEATDKEL VEHIEKILEELKKQSTD (SEQ ID NO: 428)
DHR15	NDERQKQREEVRKLAELASKATDEELIKEIKKCAQLAEELASRSTND ELIKQILEVAKLAFELASKATDEELIKEILKCCQLAFELASRSTNDELK QILEVAKLAFELASKATDEELIKEILKCCQLAFELASRSTNDEEIKQILE TAKEAFERASKATDEEEIKEILKKCQEKFEKKSRSTN (SEQ ID NO: 429)
DHR16	NDKAKEAEELLRKALEKAKEKENDETAIRCVELLKEALERAKKNNNDK AIEAVELLAKALEKALKENDETAIRCVCLLAEALLRALKNNNDKAIEA VELLAKALEKALKENDETAIRCVCLLAEALLRALKNNNDKAIEEVER LAKELEKALKENDETKIREVCERAEEELLRLKNNN (SEQ ID NO: 430)
DHR17	SSEDAREKIEQLCREAKEIAERAKQONSQEEAREAEIKLLRIAKRIAE AKQANQSEVAREAEICLCRIAKLIAELAKQANSQEVAREAEIALLRIAK LIAELAKQANQSEVAREAEICLCRIAKLIAELAKQANSQEVAREAEIAL LRIAKLIAELAKQANQSEVAREAEICLSRIAKLIEELAKQANSQEVKRE AQEALDRIQKLIEELQKQANQ (SEQ ID NO: 431)
DHR18	DIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKLASQAEAV KLACELAQEHPNADIAKLCKAASEAAEAASKAAELAQRHPDSQAAR DAIKLASQAAEAVKLACELAQEHPNADIAKLCKAASEAAEAASKAA ELAQRHPDSQAARDAIKLASQAAEAVKLACELAQEHPNADIAKKCK AASEAAEEASKAAEEAQRHPDSQKARDEIKEASQKAEVKECERACQ EHPNA (SEQ ID NO: 432)
DHR19	DEIEKVREEAEKLKKKTDDDEDVLEVAREAIRAAKEATSDEILKVIKEA LKLAKKTTDKDVLVAREAIRAAEEATDDEILKVIKEALKLAKKTTD KDVLEVAREAIRAAEEATDDEILKEIKEALKKAKETTTDTEELEKAREQI

	RKAEESTD (SEQ ID NO: 433)
DHR20	SDIEEIRQLAEELRKKSDNEEVVRKLAQEAAELAKRSTDSDVLEIVKDA LELAKQSTNEEVIKLALKA AVLAAKSTDSDVLEIVKDALELAKQSTNE EVIKLALKA AVLAAKSTDDEEVLEEVEALRRAKESTDEEEIKEELRKA VEEAESTD (SEQ ID NO: 434)
DHR21	SEKEKVEELAQRIREQLPDTLAREAQELADEARKSDDSEALKVVYL ALRIVQQLPDTLAREALELAKEAVKSTDSEALKVVYLA LRIVQQLPD TELAREALELAKEAVKSTDQEALKSVYEALQ RVQDKPNTTEEAESLE RAKEDVKSTD (SEQ ID NO: 435)
DHR22	DDAEELRERARDLLRKNGSSEEEIKKVDEELEKIVRKADSDDAVKLA VKAAALLAENGSSAEIIVK VLEELLKIVEKADSDDAVKLAVKAAALL AENGSSAEIIVK VLEELLKIVEKADSEEEVKDAVREAAELAERGSSAE EIRKQLKDRLRKVEESDS (SEQ ID NO: 436)
DHR23	SDSEKLAKRVLKELKRRGTSDEELERMKRELEKIKSATSSDAMRLAL RVVLELVRRGTSSEILEKMMRMLIKIIQSATSSDAMRLALRVVLELV RGTSSEILEKMMRMLIKIIQSATSDDQMREALRQVLEEVKRGTSSEQL ERSMRKLIKEIKRTS (SEQ ID NO: 437)
DHR24	SEAEELARRAAKEAKELCKRSTDEELCKELKKLAELLKELAE RYPDSE AAKLALKAALAEIELCKQSTDEELCEELVKLAQKLIELAKRYPDSEAA KLALKAALAEIELCKQSTDEELCEELVKLAQKLIELAKRYPDSEAAKR ALKEAKELIEQCKESTDEDECRELVKRAEELIREAKENPD (SEQ ID NO: 438)
DHR25	DERDKVRELIDRVEKELKREGTSEELIEEIRKVLKKAKEAADSDDDDEAI KVAKEIVRVILELVREGTSSSELIEEILKVLSLAAEAAKSTDDEAIK VAK EIVRVILELVREGTSSSELIEEILKVLSLAAEAAKSTDDEAAIKKAKEIVRR LELTREGTSEEEIREELKELRKKAKAKSPE (SEQ ID NO: 439)
DHR26	DECERLRQEVEKAEKELEKLAKQSTDDEEVQIAREVAKQLRRLAEEA CRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARLIRL AQEACRSNDDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVAR ELIRLAQFACRSNDDECLREASEVVKAVQELVKEAEKSTDEEEIRELLQ RAEERIREAQERCREGD (SEQ ID NO: 440)
DHR27	TRQKEQLDEVLEEIQRLAEFEARKLMTDEEEAKKIQEEAERAKEMLR AVEKVTDNEVIEKLLEVVKELRLAEEAMKKMTDEEEAAKIAKEALEA IKMLARAVEEVTNEVIEKLLEVVKELRLAEEAMKKMTDEEEAAKIA KEALEAIKMLARAVEEVTDKERIEQLREVKEEIRRAEEESRKETDDE EAAKRAREALRRIRERAREVEEDKS (SEQ ID NO: 441)
DHR28	DEEVQRIRIEEVRRAIEEVRESLERNDSEAEELAREALERVAEEVKESI KERPDRLAIEAIRALVRLAIEIVRLALEQNDSELAREVAEEALRAVAE VVKEAIRQRGDRDLAIEAIRALVRLAIEIVRLALEQNDSELAREVAEEA LRAVAEEVVKEAIRQRGDRDLAIEAIRALRRLAEEIRRLAEEQNDEL

	REVEELAREAIEEVKELERQRPGR (SEQ ID NO: 442)
DHR29	SEVEESAQVEKRAQEVREEAERRGTSQEVLDKIRVVDEARQLAQR AKESDDSEVAESALQVVREALKVVLALERTGTSEEVLKEILRVVSEAI KLALAEIKSSDSEVAESALQVVREALKVVLALERTGTSEEVLKEILRV VSEAIKLALAEIKSSDSETARRALEKVRESLKEVLEQLERTGTSEEELRE SLREVSSENIRKALEEIKSPD (SEQ ID NO: 443)
DHR30	STVKELLDRAELMRELAERASEQGSDEEEARKLLEDLEQLVQEIRRE LEETGTSSEVIRLIAKAIMLMAELALRAAEQGSDAEEAMKLLKDLLRL VLEILRELRETGTDSEVIRLIAKAIMLMAELALRAAEQGSDAEEAMKL LKDLLRLVLEILRELRETGTDKEIRKVAEEIMRRAKTALDEARQGS DAEEAMKRLKEQLRRILERLREEREKGT (SEQ ID NO: 444)
DHR31	DSYTERARKAVKRYVKEEGGSEEEAEREAEKVREEIRKKASDSYLIQA AAAVVAYVIEEGGSPEEAVKIAEEVVRRIKEKADDSYLIQAAAAVVA YVIEEGGSPEEAVKIAEEVVRRIKEKADDRELIRRAAERVAEVIERRGS PEEAVKEAEKEVKKQKEESD (SEQ ID NO: 445)
DHR32	SIQEKAKQSVIRKVKEEGGSEEEARERAKEVEERLKKEADDSTLVRAA AAVVLVLEKGGSTEEAVQRRAREVIERLKKEASDSTLVRAAAAVVLY VLEKGGSTEEAVQRRAREVIERLKKEASDEELIREAAKEVLKVLEEGGS VEEAVERARERIEELQKRSDD (SEQ ID NO: 446)
DHR33	SETEEVKKLVEEKVKKEGGSPPEAKETAKEVTEELKEESQDSTLLKVA ALVASAVLKEGGSPPEAAETAKEVVKELRKASDSTLLKVAALVASA VLKEGGSPPEAAETAKEVVKELRKASDEELLKEAARQAEEESLRQ GK SPPEAAEEAKKEVKKLKEKSQD (SEQ ID NO: 447)
DHR34	SETEEVKKLCEEKVKEGGSPPEAKETAKEVTEELKEESQDSTLLKVA ALCASAVLKEGGSCPEAAETAKEVVKELRKASDSTLLKVAALCASA VLKEGGSCPEAAETAKEVVKELRKASDEELLKEAARQAEEESLRQ GK SCPEAAEEAKKEVKKLKEKSQD (SEQ ID NO: 448)
DHR35	SEDEVAQASRYAKEQGGDPEKSREEAEKALEEVKKQATSSEALQV ALEAARYASEEGEDPAEALKEAARALEEVRRSATSSEALQVALEAAR YASEEGEDPAEALKEAARALEEVRRSATSEEDLKEALDRAREASERG QNPAESLKEAAEELKKKKEKSSD (SEQ ID NO: 449)
DHR36	SDLEKALKRFVKEEKKKGRNPPEAKKEAKKKKKLKKKSAGSSDLLTA LAKFVLEEVKGRNPPEAVKEAIKLAELKRSAGSSDLLTALAKFVLE EVRKGRNPPEAVKEAIKLAELKRSAGSSEQLEKLATKVLKEEVKKGR NPKRAVEEAIKQAKEDRKRSNS (SEQ ID NO: 450)
DHR37	SSTERAAQSVKKYLQQQKDPDQAQKKAQEVKENIEKEANSSSVIRA AAAVVFYLLQGYDPDQALKKAQEVARNIENEANSSSVIRAAAAVVF YLLQGYDPDQALKKAQEVARNIENEANSDDVIKEAAKVVKRLEE GQDPDKALEEARKRAQKTEKTTTS (SEQ ID NO: 451)
DHR38	SSTERAAQSCKYLQQQKDPDQAQKKAQEVKENIEKEANSSSVIRA

	AAACVFYLLEQGYDCDQALKKAQEVARNIENEANSSSVIRAAAACVF YLLEQGYDCDQALKKAQEVARNIENEANSDDVIKEAAKVVYKRLEE GQDCDKALEEARKRAQKTEKKTTS (SEQ ID NO: 452)
DHR39	SDLQEVADRIVEQLKREGSPPEARKEARRLIEEIKQSAGGDSSELIEVA VRIVKELEEQGRSPSEAAKEAVELIERIRRAAGGDSSELIEVAVRIVKEL EEQGRSPSEAAKEAVELIERIRRAAGGDSDRICKAVELVRELEERGRSP SEAARRAVEEIQRSVEEDGGN (SEQ ID NO: 453)
DHR40	SESDEVAKRISKEAKKEGRSEEEVKELVERFREAIKLEQGDSEAIRV AVEIADEALREGLSPEEVVELVERFVQAIQKLQENGSEAIRVAVEIAD EALREGLSPEEVVELVERFVQAIQKLQENGEEDEIQKAVETAQEQLLEE GRSPKEVVETVEEQVKEVEEKQKQGE (SEQ ID NO: 454)
DHR41	SDIEKAKRIADRAIDVVRKAAEKEGGSPEKIREALQAKRCAEKLIRL VKEAQESNSSDVREAARVALEAVRVVVRAAEEKGGSPEEVVEAVCR AVRCAEKLIRLVKRAEESNSSDVREAARVALEAVRVVVRAAEEKGGS PEEVVEAVCRAVRCAEKLIRLVKRAEESNSENVRRESARRALEKVLKT VQQAEEECKSPPEEVVEQVCRSVRKAEEQIRETQERERSTS (SEQ ID NO: 455)
DHR42	SDAEEVKKQAEELANRAYKTAQKQGESDSRAKKAELVRKAAEKL RLIERAQKEGSDALEVARQALEIARRAFETAKKQGHSAEAAAFV DVVEAAISLAELIISAKRQGDSDALEVARQALEIARRAFETAKKQGH ATEAAAFVDVVEAAISLAELIISAKRQGDQKALEIARKALQKAKENF EEAQKRGESATQAAKRFVDTVEKEIKKAQEQIKRERKGD (SEQ ID NO: 456)
DHR43	SKEEELIEKARRVAKEAIEEAKRQGKDPSEAKKAAEKLKAVEEAVKE AKRLKEEGNSELAEISEAIQVAVEAVEEAVRQKDPFKAAEAAELI RAVVEAVKEAERLKREGNSELAEISEAIQVAVEAVEEAVRQKDPF KAAEAAELIRAVVEAVKEAERLKREGNSELAKKINDTIREAVREVQ QAVEDGKDPFEAAREAAEKIRESVERVREEEKKRRGN (SEQ ID NO: 457)
DHR44	SNEQEKKDLKKAEEAAKSPDPPELIREAIERAEESGSNKAKEIILRAAEE AAKSPDPPELIRLAIEAAERSGSNKAKEIILRAAEEAAKSPDPPELIRLAIE AAERSGSEKAKEIILKRAAEEAQKSPDPPELQKLAKERERLG (SEQ ID NO: 458)
DHR45	SSEEELEKDAREASESGADPEWLREIVDLARESGDSEVIELAKRALEA AKSGADPEWLLRIVRQAEESGSSEVIELAKRALEAAKSGADPEWLLRI VRQAEESGSEVIELAKRALEAAKKGKDPKELLEVRKREESG (SEQ ID NO: 459)
DHR46	STKEEKERIERIEKEVRSPPDENIREAVRKAEEELLRENPSAEELLRRA IEAAVRAPDPEAIREAVRAAEELLRENPSAEELLRRAIEAAVRAPDP EAIREAVRAAEELLRENPSAEAKELLRRAIESAKKAPDPEAQREAKRA

	EEELRKEDP (SEQ ID NO: 460)
DHR47	STKEEKERIERIEKEVRSPOCENIREAVRKAEEELLRENPSAEELLRRA IEAAVRCPDCEAIREAVRAAEELLRENPSAEELLRRAIEAAVRCPDC EAIREAVRAAEELLRENPSAEAKELLRRAIESAKKCPDPEAQREAKRA EEELRKEDP (SEQ ID NO: 461)
DHR48	NSREEEAKRIVKEAKKSGFDPEEVEKALREVIRVAEETGNSEALKEA LKIVEEAAKSGYDPAEVAKALAEVIRVAEETGNSEALKEALKIVEEAA KSGYDPAEVAKALAEVIRVAEETGNPEELKEALKRVLEAAKRGEDPA QVAKELAEIIRRNQEEG (SEQ ID NO: 462)
DHR49	DSEEEQERIRRLKEARKSGTEESLRQAIEDVAQLAKKSQDSEVLEEAI RVILRIAKESGSEELRQAIRAVAEIAKEAQDSEVLEEAIIRVILRIAKES GSEELRQAIRAVAEIAKEAQDPRVLEEAIIRVIRQIAEESGSEARRQA ERAEIIRRRQA (SEQ ID NO: 463)
DHR50	DPEEVRREREATEEYRKNPGSDEAREQLKEAVERAEBAARSPDPEA VQVAVEAATQIYENTPGSEEAKELEIAVRAAENAAARLPDPEAVQVA VEAATQIYENTPGSEEAKELEIAVRAAENAAARLPDPEAVRVAEEAA DQIRKNTPGSELAKRADEIKKRARELLERLP (SEQ ID NO: 464)
DHR51	QSEDRKEKIRELERKARENTGSDEARQAVKEIARIAKEALEEGNADTA KEAIQRLEDLARDYSGSDVASLAVKAIKIAETALRNGYADTAKEAIQ RLEDLARDYSGSDVASLAVKAIKIAETALRNGYKETAEAAIKRLREL AEDYKGSEVAKLAEEAIERIEKVSRRER (SEQ ID NO: 465)
DHR52	QCEDRKEKIRELERKARENTGSDEARQAVKEIARIAKEALEEGCCDTA KEAIQRLEDLARDYSGSDVASLAVKAIKIAETALRNGCCDTAKEAIQ RLEDLARDYSGSDVASLAVKAIKIAETALRNGCKETAEEAIKRLREL AEDYKGSEVAKLAEEAIERIEKVSRRER (SEQ ID NO: 466)
DHR53	SNDEKEKLKELLKRAEELAKSPDPEDLKEAVRLAEVVRERPGSNLA KKALEIILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSNLAKKALE IILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSELAKKALEIIERAA EELKKSPDPEAQKEAKKAEQKVRERPG (SEQ ID NO: 467)
DHR54	TTEDERRELEKVARKAIEAAREGNTDEVREQLQRALEIARESQTTEAV KLALVVARVAIEAARRGNTDAVREALEVALEIARESQTTEAVKLAL EVVARVAIEAARRGNTDAVREALEVALEIARESQTTEAVRLALEVVK RVSDAEKKQGNEDAVKEAEVVRKKIEEESG (SEQ ID NO: 468)
DHR55	SSVAEEIEKRAKKISKELKKEGKNPEWIEELQRAADKLVEVARRATSS DALEIAKRAVKIAEELAKQGSNPKWIAELLKAAAKLVEVAARATSSD ALEIAKRAVKIAEELAKQGSNPKWIAELLKAAAKLVEVAARATSPKA LKQAKEAVKEAEELAKKGRNPKEIAEELKKRAKEVEKLARST (SEQ ID NO: 469)
DHR56	SSVAEEIEKRCCKISKELKKEGKNPEWIEELQACDKLVEVARRATSS DALEIAKRCVKIAEELAKQGSNPKWIAELLKACAKLVEVAARATSSD

	ALEIAKRCVKIAEELAKQGSNPKWIAELLKACAKLVEVAARATSPKA LKQAKECVKEAEELAKKGRNPKEIAEELKKCAKEVEKLARST (SEQ ID NO: 470)
DHR57	STEELKKVLERVRELSERAKESTDPEEALKIAKEVIELALKAVKEDPST DALRAVLEAVRLASEVAKRVTDPKALKIAKL VIELALEAVKEDPST DALRAVLEAVRLASEVAKRVTDPKALKIAKL VIELALEAVKEDPSEE AKRAVEEAKRLAEEVSKRVTDPELSEKIRQLVKELEEEEAQKEDP (SEQ ID NO: 471)
DHR58	STEELKKVLERVRELCERAKESTDPEEALKIAKEVIELALKAVKEDPST DALRAVLEAVRCACEVAKRVTDPKALKIAKL VIELALEAVKEDPST DALRAVLEAVRCACEVAKRVTDPKALKIAKL VIELALEAVKEDPSE EAKRAVEEAKRCAEEVSKRVTDPELSEKIRQLVKELEEEEAQKEDP (SEQ ID NO: 472)
DHR59	KTEVEKKAKEVIKEAKELAKELDSEEAKKVVERIKEAAEA AKRAAEQ GKTEVAKLALKVLEEAIELAKENRSEEALKVVLEIARAALAAAQAAE EGKTEVAKLALKVLEEAIELAKENRSEEALKVVLEIARAALAAAQAA EEGKSDEARDALRRLEEAIEEAKENRSKESLEKVREEAKEAEQQAED AREG (SEQ ID NO: 473)
DHR60	TDIKKKAEEIIEAKKQGSSEDAIRLAQEAKKQGTDILVRAAEIVVRAQ EQGSSEDAIRLAKEASREGTDILVRAAEIVVRAQE QGSSEDAIRLAKEAS REGTPTLVKAAEKVVRAQQKGSQDTIEKAKEESREG (SEQ ID NO: 474)
DHR61	TDIKKKAEEIIEAKKQGSSEDAIRLAQECKKQGTDICVRAAEIVVRAQ EQGSSEDAIRLAKECSRREGTDICVRAAEIVVRAQE QGSSEDAIRLAKECSR EGTPTCVKAAEKVVRAQQKGSQDTIEKAKEESREG (SEQ ID NO: 475)
DHR62	DNDEKRRKRAEKALQRAQAEKKGDVEEAVRAAQEAVRAAKESGDN DVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDNDV LRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDQDVL KVSEQAERISKEAKKQGNSEVSEEARKVADEAKKQTG (SEQ ID NO: 476)
DHR63	DPDEDRERLKEELKKIREALREAKEKPDPEEIKRALREVLEAIRRILKL AERAGDPLAREALKEINKVIREALEIAKRVDPDEVIKEALRVVLEAIR AILKLAEQAGDPLAREALKEINKVIREALEIAKRVDPDEVIKEALRVV LEAIRAILKLAEQAGDPLAREALEEIDKVIDEAQEISERVPDEEVQRE AQEVIKEADRARKKLSEQSG (SEQ ID NO: 477)
DHR64	DPEDELKRVEKL VKEAEELLRQAKEKGSEEDLEKALRTAEAAAREAK KVLEQAEKEGDPEVALRAVELVVRVAELLRIAKESGSEEALERALR VAEEAARLAKRVLELAEKQGDPEVALRAVELVVRVAELLRIAKESG SEEALERALRVAEEAARLAKRVLELAEKQGDPEVARRAVELVKRVAE LLERIAESGSEEAKERAERVREEARELQERVKELREREG (SEQ ID

	NO: 478)
DHR65	DPEDELKRVEKLVKEAEELLRQCKEKGSEECLEKALRTAEEAAREAK KVLEQAEKEGDPEVALRAVELVVRVAELLRLICKESGSEECLEALRV AEEAARLAKRVLELAEKQGDPEVALRAVELVVRVAELLRLICKESGS EECLERALRVAEEAARLAKRVLELAEKQGDPEVARRAVELVKRVAEL LERICRESGSECKERAERVREEARELQERVKELREREG (SEQ ID NO: 479)
DHR66	TSDDDKVREAEERVREAIERIQRALKKRDTDPDARKALEAAKLLKV EKAKKRGTSDAIKVAEAAARVAEAIARILEALNERDTPDARKALRAAI KLAENVVYKAAESGTSDAIKVAEAAARVAEAIARILEALNERDTPDAR KALRAAIKLAENVVYKAAESGTTEALKVAEKAARVAEKIARILEKLN RDTPEARKKLQAIKEAEKVKESEQ (SEQ ID NO: 480)
DHR67	TSEIDKLIKLRQTAKEVKREAEERKRRSTDPTVREVIERLAQLALDV AEEAARLIKKATTSEVAKLVWKLARTAEVIREAIERAERSTDPEVIRV ILELARLAAEVAKEAARLIVKATTSEVAKLVWKLARTAEVIREAIERA ERSTDPEVIRVILELARLAAEVAKEAARLIVKATTEEVAKKVWKEAYR AIEEIRKAIEKAERSTDPNEIKKILEEARKKAAEAIERAKEIVKST (SEQ ID NO: 481)
DHR68	TPRERLEEAKERVEEIRELIDKARKLQEQGNKEEAKEVLREAREQIRE VTRELEEIKNSTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEA KVLREATELIKRVTELLEKIAKNSTPELALRAAELLVRLIKLLIEIAKL LQEQGNKEEAKEVLREATELIKRVTELLEKIAKNSTPELAKRAAELL KRLIELLKEIAKLLEEEGNEDEAEKVKEEAKLEERVRELEERIRKNSD (SEQ ID NO: 482)
DHR69	NPQEDLERAKEVVRSVEEVLQRAKEAQREGDKEKVERLIKEAENQIR KARELLERVVRQNPDDPEVLLRVAELIVRLVEVVLELAKLAEKNGDK EQVERLIQTAEELIREARELLERSREIPDNPEVLLRVAELIVRLVEVVL ELAKLAEKNGDKEQVERLIQTAEELIREARELLERSREIPDNPESLKR VAELIKRLVKVDEL SKLAERNGDRDQVERLRQLAEELRREAEELEE RVRRERPD (SEQ ID NO: 483)
DHR70	STEEKIEEARQSIKEAERSLREGNPEKAREDVRRALVRELEKLARKT GSTEVLEIAARLAIEVARVALKVGPETAREAVRTALELVQELERQAR KTGSTEVLEIAARLAIEVARVALKVGPETAREAVRTALELVQELERQ ARKTGSDEVLKRAELAKEVARVAKEVGPETARQARETAERLREEL RRNREKKG (SEQ ID NO: 484)
DHR71	DPEEILERAKESSLERAREASERGDEEFFRKAEEKALELAKRLVEQAKK EGDPPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVE VASKEGDPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAK RLVEVASKEGDPELVLEAAKVAEEVRKLAKKQGDEEVYEKARETAR EVKEELKRVREEKG (SEQ ID NO: 485)

DHR72	DSTKEKARQLAEEAKETAKEKVGDPPELIKLAEQASQEGDSEKAKAAILLA AEAARVAKEVGDPELIKLALEAARRGDSEKAKAAILLAEEAARVAKEV GDPELIKLALEAARRGDSEKARAILEAAERAREAKERGDPEQIKKARE LAKRG (SEQ ID NO: 486)
DHR73	DAEEEEAKEAKRAQEAIELARKGNPEEARKVAEEFARERAERVREEAE KRGDAEVLALVAIALALVAIALAEVGNPEEAREVAERAKEIAERVREL AEKRGDAEVLALVAIALALVAIALAEVGNPEEAREVAERAKEIAERV RELAEKRGDARVLKLVAKALELVAEALKKVGNPPEEAREVEERAREIK ERVRRLLLEEK (SEQ ID NO: 487)
DHR74	DSEADRIKKLQKEIKEVEQEARDSNDDEERELLKRLAEALKRAAEAV KRAQESGDSEAIRIUKKLVEITEVVREARKSTDKEEIELLIRLAELAR AAEAVADAASKSGDSEAIRIUKKLVEITEVVREARKSTDKEEIELLIRL AEALARAEEAVADAASKSGDQEAIKRIKKLVKKIIEVVRKARKSTNKK EIEKLIRKAEKLARKAEQIAEDAKRG (SEQ ID NO: 488)
DHR75	DSEKEKATELAERAQDVASRVEEEARREGSRELIEIARELRERAEAS QEGDSEKAKAILLAAKAVLVAVEVYERAKRQGSDELREIARELAKEA LRAAQEGDSEKAKAILLAAKAVLVAVEVYERAKRQGSDELREIAREL AKEALRAAQEGDSEKARAILEAAAREVLRAVEQYERAKRRGDDDERE RAREEAREALERAREG (SEQ ID NO: 489)
DHR76	NPELEEWIRRAKEVAKVEKVAQRAEEEGNPDLRDSAKELRRAVEEA IEEAKKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALE LVRAVIEAIEEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNR DLFRAALELVRAVIEAIEEAVKQGNPELVERVARLAKKAAELIKRAIR AEKEGNRDERREALERVREVIERIEELVRQG (SEQ ID NO: 490)
DHR77	NSDEEEAREWAERAEEAAKEALEQAKREGDEDARRVAEELEKQAE ARRKKDSEEAEEAVYWAARAVLAALAELEQAKREGDEDARRVAEELL RQAEAAARKKNSEEAEEAVYWAARAVLAALAELEQAKREGDEDARR VAEELLRQAEAAARKKNPEEARAVYEAARDVLEALQRLLEEAKRRGD EEERREAEERLRQAEERARKK (SEQ ID NO: 491)
DHR78	NSDEEEAREWAERAEEAAKEALEQAKREGDEDARRCAEELEKQAE ARRKKDSEEAEEAVYWAARAVLAALAELEQAKREGDEDARRCAEELL RQACEAARKKNSEEAEEAVYWAARAVLAALAELEQAKREGDEDARR CAEELLRQACEAARKKNPEEARAVYEAARDVLEALQRLLEEAKRRGD EEERREAEERLRQACERARKK (SEQ ID NO: 492)
DHR79	SSDEEEARELIERAKEAAERAQEAERTGDPRVRELARELKRLAQEAA EEVKRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELAREL VRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPE VRELARELVRLAVEAAEEVQRNPSSSEVNEALKKIVKAIQEAVESLRE AEESGDPEKREKARERVREAVERAEEVQRDPS (SEQ ID NO: 493)
DHR80	NSEELERESEEAERRLQEARKRSEERARERGDLEALAEALIEEARAVQE

	LARVASERGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEV LALA LIAIALAVLALAEVASSRGNSEEAERASEKAQRVLEEARKVSEEAREQ GDDEV LALALIAIALAVLALAEVASSRGNKEEAERAYEDARRVEEEA RKVKESAEEQGDSEVKRLAEAEQLAREARRHVQETRG (SEQ ID NO: 494)
DHR81	NSEELERESEEAERRLQEARKRSEERERGD LKELAEALIEEARAVQE LARVACERGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEV LALA LIAIALAVLALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQ GDDEV LALALIAIALAVLALAEVACCRGNKEEAERAYEDARRVEEEA RKVKESAEEQGDSEVKRLAEAEQLAREARRHVQECRG (SEQ ID NO: 495)
DHR82	NDEEVQEAVERAEEELREEAEELIKKARKTGDPELLRKALEALEEAVR AVEEAIKRNPNDNAVETAVRLARELKKVAEELQERAKKTGDPELLK LALRALEVAVRAVELAIKSNPDNAVETAVRLARELKKVAEELQER AKKTGDPELLKLALRALEVAVRAVELAIKSNPDNEEAVETAKRLAEE LRKVAELLEERAKETGDPELQELAKRAKEVADRARELAKKSNPN (SEQ ID NO: 496)
DHR83	NDEEVQEACERAEEELREEAEELIKKARKTGDPELLRKALEALEEAVRA VEEAIKRNPNDNAVETACRLARELKKVAEELQERAKKTGDPELLKL ALRALEVAVRAVELAIKSNPDNAVETACRLARELKKVAEELQERA KKTGDPELLKLALRALEVAVRAVELAIKSNPDNEECVETAKRLAEE RKVAELLEERAKETGDPELQELAKRAKEVADRARELAKKSNPN (SEQ ID NO: 497)

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. The polypeptides described herein may be chemically synthesized or recombinantly expressed. The polypeptides may be linked to other compounds to promote an increased half-life *in vivo*, such as by PEGylation, HESylation, PASylation, glycosylation, or may be produced as an Fc-fusion or in deimmunized variants. Such linkage can be covalent or non-covalent as is understood by those of skill in the art.

As will be understood by those of skill in the art, the polypeptides of the invention may include additional residues at the N-terminus, C-terminus, or both that are not present in the polypeptides of Tables 1-2; these additional residues are not included in determining the percent identity of the polypeptides of the invention relative to the reference polypeptide.

In one embodiment, the polypeptide comprises at least one conservative amino acid substitution. As used herein, "conservative amino acid substitution" means amino acid or nucleic acid substitutions that do not alter or substantially alter polypeptide or polynucleotide function or other characteristics. A given amino acid can be replaced by a residue having similar physiochemical characteristics, e.g., substituting one aliphatic residue for another (such as Ile, Val, Leu, or Ala for one another), or substitution of one polar residue for another (such as between Lys and Arg; Glu and Asp; or Gln and Asn). Other such conservative substitutions, e.g., substitutions of entire regions having similar hydrophobicity characteristics, are well known. Polypeptides comprising conservative amino acid substitutions can be tested in any one of the assays described herein to confirm that a desired activity, e.g. antigen-binding activity and specificity of a native or reference polypeptide is retained.

Amino acids can be grouped according to similarities in the properties of their side chains (in A. L. Lehninger, in *Biochemistry*, second ed., pp. 73-75, Worth Publishers, New York (1975)): (1) non-polar: Ala (A), Val (V), Leu (L), Ile (I), Pro (P), Phe (F), Trp (W), Met (M); (2) uncharged polar: Gly (G), Ser (S), Thr (T), Cys (C), Tyr (Y), Asn (N), Gln (Q); (3) acidic: Asp (D), Glu (E); (4) basic: Lys (K), Arg (R), His (H). Alternatively, naturally occurring residues can be divided into groups based on common side-chain properties: (1) hydrophobic: Norleucine, Met, Ala, Val, Leu, Ile; (2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln; (3) acidic: Asp, Glu; (4) basic: His, Lys, Arg; (5) residues that influence chain orientation: Gly, Pro; (6) aromatic: Trp, Tyr, Phe. Non-conservative substitutions will entail exchanging a member of one of these classes for another class. Particular conservative substitutions include, for example; Ala into Gly or into Ser; Arg into Lys; Asn into Gln or into His; Asp into Glu; Cys into Ser; Gln into Asn; Glu into Asp; Gly into Ala or into Pro; His into Asn or into Gln; Ile into Leu or into Val; Leu into Ile or into Val; Lys into Arg, into Gln or into Glu; Met into Leu, into Tyr or into Ile; Phe into Met, into Leu or into Tyr; Ser into Thr; Thr into Ser; Trp into Tyr, Tyr into Trp; and/or Phe into Val, into Ile or into Leu. As noted above, the polypeptides of the invention may include additional residues at the N-terminus, C-terminus, or both. Such residues may be any residues suitable for an intended use, including but not limited to detection tags (i.e.: fluorescent proteins, antibody epitope tags, etc.), linkers, ligands suitable for purposes of purification (His tags, etc.), and peptide domains that add functionality to the polypeptides.

In another embodiment, the invention provides protein assemblies, comprising a plurality of polypeptides of the present invention having the same amino acid sequence. As

disclosed herein, the polypeptides of the invention represent novel repeat proteins with precisely specified geometries, and thus self-assemble into the protein assemblies of the invention.

In a further aspect, the present invention provides isolated nucleic acids encoding a polypeptide 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 poly A 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 polypeptides of the invention.

In another aspect, the present invention provides recombinant expression vectors comprising the isolated nucleic acid of any aspect 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" operably 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 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 host cells is 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, TX). The expression vector must be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. In various embodiments, the expression vector may comprise a plasmid, viral-based
5 vector, or any other suitable expression vector.

In a further aspect, the present invention provides host cells that comprise the recombinant expression vectors disclosed herein, wherein the host cells can be either prokaryotic or eukaryotic. The cells can be transiently or stably engineered to incorporate the expression vector of the invention, using standard techniques in the art, including but not
10 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, NY). A
15 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. The expressed polypeptide can be recovered from the cell free extract, but preferably they are recovered from the culture
20 medium. Methods to recover polypeptide from cell free extracts or culture medium are well known to the person skilled in the art.

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
25 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.

30 EXAMPLES

In repeat proteins, the interactions between adjacent units define the shape and curvature of the overall structure⁶. While in nature the sequences of these units generally differ, highly stable repeat proteins with identical units^{7,8} have been designed for several

families⁹⁻²¹ and, for leucine rich repeats, customized designed units allow control of curvature²² and new architectures¹⁷. All designed repeat structures to date have been based on naturally occurring repeat protein families. These families may cover all stable repeat protein structures that can be built from the 20 amino acids or, alternatively, natural evolution may
5 only have sampled a subset of what is possible.

To explore the range of possible repeat protein structures, we generated new repeat protein backbone arrangements and designed sequences predicted to fold into these structures (Fig. 1). Our designs are entirely *de novo*; they are not based on naturally occurring repeat proteins. We focused on helix-loop-helix-loop as the basic repeating unit, as this is the
10 simplest unit from which a wide diversity of curvatures can be generated (the simpler single helix-loop unit generates only straight rod-like models). The lengths of the two helices were varied between 10 and 28 residues, and the lengths of the two turns, from 1 to 4 residues. Starting conformations for four tandem repeats of each of the 5776 (19×19×4×4) combinations of helix and loop lengths were generated by setting the backbone torsion angles
15 to ideal helix values for helices and extended chain values for loops. Rosetta Monte Carlo fragment assembly²³ was carried out to generate compact structures; each Monte Carlo move was made at the equivalent position in each repeat to preserve symmetry²⁰. Rosetta design calculations²⁴ were then used to identify low energy amino acid sequences with good core packing²⁵. At each step in the Monte Carlo - simulated annealing design process, a position is
20 picked at random, and the current residue is replaced by a randomly selected amino acid and side chain conformation (rotamer); a detailed all-atom energy function is then evaluated. Identical substitutions were carried out in each copy at each move to maintain sequence identity between the four repeats; exposed hydrophobic residues in the N- and C-terminal repeats were switched to polar residues in a second round of sequence design, generating
25 specialized capping repeats. All steps in the design process were completely automated, and the calculations were carried out without manual intervention. Designs with low energies and complementary core side chain packing were identified, and for the amino acid sequence of each of these designs, multiple independent Rosetta *de novo* folding trajectories²⁶ were carried out starting from an extended chain. The structures and energies of the sampled
30 conformations map out an energy landscape for each protein (Fig. 5).

Designed helical repeat proteins (DHRs), for which the design model had much lower energy than any other conformations sampled in the *de novo* folding trajectories, were selected and found to span a wide array of architectures. As the rigid body transform relating adjacent repeat units is identical throughout each design by construction, and since the

repeated application to an object of an identical rigid body transformation produces a helical array, the designs all have an overall helical structure⁶. It is thus convenient to classify these architectures based on three parameters defining a helix²²: the radius (r), the twist between adjacent repeats around the helical axis (ω) and the translation between adjacent repeats
 5 along the helical axis (z). Because the repeat units are connected and form well packed structures, the three parameters are coupled. The arc length in the x-y plane spanned by a repeat unit is $\sim r\omega$, and the total length of a unit is $\sim \sqrt{(r\omega)^2 + z^2}$, hence the radius(r)-twist(ω) distribution has a hyperbolic shape with highly twisted structures having a smaller radius. Models with high r and high ω do not form a continuous protein core and are
 10 discarded during the backbone generation. Similarly, low energy structures do not have high (>16 Å) z values as helices in adjacent repeats cannot then closely pack. Despite these geometric constraints, the wide range of helical parameters observed in the design models highlights the high level of complexity that can be generated even for a pair of helices. In contrast, native helical repeat proteins span a much narrower range of helical parameters with
 15 very few straight (high r , low ω) or highly twisted (low r , high ω) geometries.

We selected for experimental characterization 83 designs spanning the range of α -helix and loop lengths and overall helical architectures; 26 of these contain disulphide bonds. For each of the designs, we obtained a synthetic gene encoding an N-terminal capping repeat, two internal repeats, and a C-terminal capping repeat including a 6-histidine tag. The proteins
 20 were expressed in *Escherichia coli* and purified by affinity chromatography. 74 of the 83 designs were expressed solubly and had the expected alpha helical CD spectrum at 25 °C, and 72 were stably folded at 95 °C. 55 of these (66% of the original experimental set) were predominantly monomeric by analytical size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS); DHR49 and DHR76 were dimeric in solution. This
 25 group had the same fraction of proteins with disulphide bonds as the initial set (Fig. 2a), indicating that disulphide bonds did not provide any particular advantage in expression, solubility, or folding efficiency by further stabilizing the fold. Representative data on six of the designs are shown in Fig. 2b.

We solved the crystal structures of 15 of the designs (Fig. 3) with resolutions between
 30 1.20 Å and 3.35 Å. The design models closely match the crystal structures with C α RMSDs from 0.7 Å to 2.5 Å and recapitulate the side chain orientations within the hydrophobic core (Figs. 3 and 6). The designed disulfide bonds are all formed in the structures of DHR4 and DHR7 but not in the structures of DHR5 and DHR18 due to slight structural shifts relative to

the design models. The accuracy of the design models was sufficiently high that all of the crystal structures but DHR5 could be solved by molecular replacement. These repeat proteins are among the largest crystallographically validated protein structures designed completely *de novo*, ranging in size from 171 residues for DHR49 to 238 residues for DHR64. The
5 crystal structures illustrate both the wide range of twist and curvature sampled by our repeat protein generation process and the accuracy with which these can be designed.

To characterize the structures for proteins that were reticent to crystallization and analyze all 55 proteins in solution, we used small angle X-ray scattering (SAXS)^{27,28}. We collected SAXS profiles for each design, and compared them to scattering profiles calculated
10 from the design models and from crystal structures. For 43 of the designs, the radius of gyration, molecular weight, and distance distributions computed from the SAXS data corresponded to those computed from the models. For DHR49 and DHR76, we used the dimer orientation in the crystal for the fitting; the crystallographically confirmed DHR5 was unsuitable for SAXS as it formed higher order species. To further assess the fit between
15 models and experimental data, we employed the volatility ratio (Vr), which is more robust to experimental noise than the traditional χ^2 comparison used in SAXS²⁹. We used the Vr values of the design models confirmed by crystallography for calibration; designs for which the Vr value between model and experimental data was less than 2.5 were considered successful. All 43 designs with radii, molecular weights, and distances consistent with the
20 SAXS data are below the Vr threshold. Furthermore, for almost all of the designs, the theoretical scattering profile computed from the design model more closely matches its own experimental scattering profile than the experimental scattering profiles of structurally dissimilar designs.

The crystallographic and SAXS data together structurally validate 44 of the 55
25 designs that were folded and monodisperse -- more than half of the 83 that were experimentally characterized. We randomly selected two designs confirmed by crystallography, two confirmed by SAXS, and two not confirmed by SAXS, and examined their guanidine hydrochloride (GuHCl) unfolding profiles. In contrast to almost all native proteins, four of the six designs do not denature at GuHCl concentrations up to 7.5 M; the
30 other two, which were confirmed by SAXS but did not yield crystals, have denaturation midpoints above 3 M (Fig. 7). Hence, even the apparent failures are well folded proteins; small amounts of association may be responsible for the discrepancies between computed and observed SAXS spectra rather than deviations from the design models.

We show here that a wide range of novel repeat proteins can be generated by tandem repeating a simple helix-loop-helix-loop building block. As illustrated by the comparison of 15 design models to the corresponding crystal structures (Fig. 3), our approach allows precise control over structural details throughout a broad range of geometries and curvatures. The design models and sequences are remarkably different from each other and from naturally occurring repeat proteins, without any significant sequence or structural homology to known proteins. This work achieves key milestones in computational protein design: the design protocol is completely automatic, the folds are unlike those in nature, more than half of the experimentally tested designs have the correct overall structure as assessed by SAXS, and the crystal structures demonstrate precise control over backbone conformation for proteins over 200 amino acids. The observed level of control over the repeating helix-loop-helix-loop architecture shows that computational protein design has matured to the point of providing alternatives to naturally occurring scaffolds, including graded and tunable variation difficult to achieve starting from existing proteins. We anticipate that the 44 successful designs described in this work, and sets generated using similar protocols for other repeat units, will be widely useful starting points for the design of new protein functions and assemblies.

Naturally occurring repeat protein families, such as ankyrins, leucine rich repeats, TAL effectors and many others, play central roles in biological systems and in current molecular engineering efforts. Our results suggest that these families are only the tip of the iceberg of what is possible for polypeptide chains: there are clearly large regions of repeat protein space that are not sampled by currently known repeat protein structures. Repeat protein structures similar to our designs may not have been characterized yet, or perhaps may simply not exist in nature.

Methods

25 Similarity search.

BLAST^{30,31} and HHSEARCH³² sequence similarity searches were performed with default settings. HHSEARCH was run on Pfam³³. Sequence alignments were depicted using Jalview³⁴. The structural similarity between designs and known helical repeat proteins was assessed by TM-align³⁵ on RepeatsDB³⁶ representative structures.

30 Protein expression and characterization.

Genes were synthesized and cloned in vector pET21 by GenScript (Piscataway, NJ). Proteins were expressed in *E. coli* BL21(DE3), induced with 250 μ M isopropyl- β -D-thiogalactopyranoside (IPTG) overnight at 22°C and purified by metal ion affinity chromatography (IMAC) and size exclusion chromatography (SEC) as described by

Parmeggiani *et al.*²⁰ Cells were lysed by sonication and the clarified lysate was loaded on a NiNTA superflow column (Qiagen). Lysis and washing buffer was Tris 50mM, pH 8, NaCl 500mM, imidazole 30mM, glycerol 5% v/v. Lysozyme (2 mg/ml), DNaseI (0.2 mg/ml) and protease inhibitor cocktail (Roche) were added to the lysis buffer before sonication. Proteins were eluted in Tris 50 mM, pH 8, NaCl 500mM, imidazole 250mM, glycerol 5% v/v and dialyzed overnight either in tris 20mM, pH 8, NaCl 150mM. Protein concentrations were determined using a NanoDrop spectrophotometer (Thermo Scientific). Except as indicated above, enzymes and chemicals were purchased from Sigma-Aldrich. Secondary structure content, thermal stability and denaturation in presence of guanidine hydrochloride (GuHCl) were monitored by Circular Dichroism using an AVIV 420 spectrometer (Aviv Biomedical, Lakewood, NJ). Thermal denaturation was followed at 220nm in Tris 20mM, 50mM NaCl, pH 8. Proteins were considered folded if they had the expected alpha helical CD spectrum at 25°C and had either a sharp transition in thermal denaturation or a loss of less than 20% of 220 nm CD signal at 95°C. Chemical denaturation was monitored in a 1cm path-length cuvette at 222nm with protein concentration of 0.05mg/ml in phosphate buffer 25mM NaCl 50mM pH 7. The GuHCl concentration was automatically controlled by a Microlab titrator (Hamilton). Oligomeric state was assessed by Analytical Gel Filtration coupled to Multiple Angle Light Scattering (AFG-MALS). A Superdex 75 10/300 GL column (or superdex200 increase for DHR59, 84, 93) (GE Healthcare) equilibrated in Tris 20mM, NaCl 150mM, pH 8 was used on a HPLC LC 1200 Series (Agilent Technologies) connected to a miniDAWN TREOS (Wyatt Technologies). Protein molecular weights were confirmed by mass spectrometry on a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). 74 of the 83 designs were expressed solubly and had the expected alpha helical CD spectrum at 25°C. 72 were stably folded at 95°C. DHR36 has T_m=75°C and DHR13 has a broad transition with T_m=62°C. Fifty-five of these were predominantly monodisperse. DHR49 and 76 were dimeric in solution.

Crystallization.

Proteins were purified using NiNTA resin and SEC on a superdex 75 column (GE healthcare). Pure fractions in the gel filtration buffer (20 mM Tris pH 8.0, 150 mM NaCl) were pooled and concentrated for crystallography. Initial crystallization trials were performed using the JCSG core I-IV screens at 22 °C, and crystals were optimized if necessary. Drops were set up with the Mosquito HTS using 100 nL protein and 100 nL of the well solution. Crystals were cryoprotected in the reservoir solution supplemented with ethylene glycol, then flash cooled and stored in liquid nitrogen until data collection. All

diffraction data were collected at the Advanced Light Source (ALS) at beamline 8.3.1 or beamline 8.2.1. Data reduction was carried out using XDS³⁷ and HKL2000 (HKL Research). Most of the structures reported here were solved by molecular replacement using Phaser. Search models were generated by *ab initio* folding of the designed sequences in Rosetta and a set of the lowest energy 10-100 models was selected for molecular replacement trials. DHR5 was the only structure which could not be readily solved by molecular replacement. However, due to the presence of 6 cysteine residues in the native protein, the DHR5 structure was solved by sulfur single wavelength anomalous dispersion (S-SAD) using a dataset collected at 7235 eV. Rigid body, restrained refinement with TLS and simulated annealing were carried out in Phenix³⁸. Manual adjustment of the model was carried out in Coot³⁹. The structures were validated using the Quality Control Check v2.8 developed by JCSG, which included Molprobity⁴⁰ (publicly available at the [smb.slac.stanford](http://smb.slac.stanford.edu) web site).

SAXS.

SAXS data on SEC-purified protein were collected at the SIBYLS 12.3.1 beamline at the Advanced Light Source, LBNL^{28,41,42}. Scattering measurements were performed on 20 microliter samples and loaded into a helium-purged sample chamber, 1.5 m from the Mar165 detector. Data were collected on both the original gel filtration fractions and samples concentrated ~2x-8x from individual fractions. Fractions prior to the void volume and concentrator eluates were used for buffer subtraction. Sequential exposures (0.5, 1, 2, and 5s) were taken at 12 keV to maximize signal to noise with visual checks for radiation-induced damage to the protein. The data used for fitting were selected for having higher signal to noise ratio and lack of radiation-induced aggregation. In case of concentration dependency, the lowest concentration was used. Models for SAXS comparison were obtained by adding the flexible C-terminal tag present in the constructs to the original designs and the crystal structures, generating 100 trajectories for each starting model by Monte Carlo fragment insertion²³. The results were clustered in Rosetta with a cluster radius of 2 Å and the cluster centers were used for comparison to the experimental data. We used FOXS^{43,44} to calculate scattering profiles from cluster centers and fit them to the experimental data. The quality of fit between models and experimental SAXS data is usually assessed by the χ value⁴⁵, which, however, suffers from over-fitting in case of noisy datasets and domination of the low region of the scattering vector (q) on the value²⁷. To avoid artificially low values that represent false positives, we instead used Volatility Ratio (Vr)²⁹ as primary metric for fit in the range of $0.015 \text{ \AA}^{-1} < q < 0.25 \text{ \AA}^{-1}$. Vr values of models with available crystal structures range from

0.7 to 2.3. $V_r=2.5$ was selected as upper threshold to consider a design as validated by SAXS.

Model profiles for V_r similarity maps were obtained with a standardized fit procedure by averaging the scattering profile of the cluster centers from the five largest clusters and
5 fitting the solvent hydration layer with parameters $C1=1.015$ and $C2=2.0$ for all the models. V_r was calculated in the range $0.04 \text{ \AA}^{-1} < q < 0.3 \text{ \AA}^{-1}$. The order of display was derived by shape similarity of original computational models using the program damsupsup⁴⁶ for superposition.

Computational protocol

10 We have developed a method for construction of Designed Helical Repeats (DHRs) depicted in Fig. 4 and described below. We designed proteins based on repeating units formed by two helices and two loops. For all proteins this design process was completely automated and no manual refinement was involved. Using this protocol 69 proteins with diverse architectures were selected from the *in silico* candidates. For 14 models, an additional
15 version that included disulphide bonds was selected, for a final list of 83 proteins that were experimentally tested. This design method has progressed over the duration of this research and only the final design method is described below. The database described in section 1 of the supplementary corresponds to the technique used to make DHR56-83. (a) For DHR1-4,9,11-18 the repeat backbone at the centroid level was symmetric, with first and second
20 helices and first and second loops having the same length and conformation. The design stage was not restricted, introducing structural and sequence variability between the two halves of the repeat. (b) A higher disulfide score threshold of 1.5 was initially used which resulted in many disulfide-containing structures being non-functional. (c) We initially used ambiguous constraints between the helices. Ambiguous constraints gave a score bonus to
25 centroid models when a helix was within 10 Å to a helix in adjacent repeat. These constraints were found to disrupt loops and result in many structures that would not fold during simulations. (d) DHR31-55 contained a displacement between helices, which resulted in highly twisted structures. This displacement was observed when the ABEGO loop types GBB and BAB were coupled with specific helix lengths. An improved sampling strategy
30 with increased number of Monte Carlo steps was also used in these cases.

In some examples, computer software such as the Rosetta software suite (or, briefly, Rosetta), can be used to carry out at least part of the herein-described methods, protocols, and/or techniques. However, the herein-described methods and techniques are not limited

to use of Rosetta or any other specific software package. For example, other software programs could be used in conjunction with this method to model multi-component symmetric protein nanostructures. As will be understood by those of skill in the art, the implementation of the design methods described herein is non-limiting, and the methods are in no way limited to the implementation disclosed herein.

Each of the following sections describes one step in Rosetta examples and corresponds to the flow chart in Fig. 4.

1 Backbone Design

The backbone design stage employs a simplified side chain representation (centroid)^{S1}. The backbone assembly procedure begins by picking fragments harvested directly from a non-redundant set of structures from PDB^{S2}. The fragments contain only residues that fall into the space of phi-psi backbone angles of either helices or loops depending on the desired secondary structure. Loop fragments could be further specified to fall within desired ABEGO bins³ as described by Koga *et al.*^{S4}.

The fragments were assembled using a Monte-Carlo sampling procedure that was initialized with ideal-helices and extended loops. After every fragment sampling step, which was allowed only in the first repeat unit and at the junction between the first and the second units, the change was propagated to all downstream repeats and scored. The score function we used considered van der Waals interactions, packing, values of backbone dihedral angles, and radius of gyration (RG) that was applied to only the first and second repeat-unit (RG-local). The RG term promotes the formation of globular proteins so applying RG to the whole model produced only highly curved structures. The sampling procedure in the database used 1500 Monte Carlo fragment insertions and was further improved to 3200 steps ordered as following: 100 Monte Carlo moves with 9 residue fragments then 100 moves with 3 residue fragments, both allowed only in loops. The loop sampling was followed by 1500 moves with 9 residue fragments and 1500 moves with 3 residue fragments, both in helices and loops (improved sampling). The improvements resulted in a 3.3 times increase of acceptance at the centroid stage. The backbone was represented as poly-tyrosine during the centroid building, maintaining enough space within the core to accommodate both small and large side chains in the design step.

Using this procedure we designed 2.88 million backbones by making 500 structures for each of 5776 different secondary structure combination.

2 Backbone quality filter: RMSD loop threshold and motif score

Designed backbones were screened for native-like features. First, loops were checked so that there was at least one 9-residue fragment from the PDB database within 0.4Å RMSD on every position in the structure (RMSD loop threshold). To do this we used the worst9mer filter in Rosetta^{S1}. Second, the design-ability of each residue was measured by the number of pairwise side chain interactions observed in the PDB database, considering the backbone position of the two residues involved (motif score, unpublished results). Backbones with fewer than 1.5 interactions per residue were filtered out. Of the 2.88 million initial backbones 66,776 structures passed these filters.

3 Sequence design - Fast

Starting from the filtered backbone conformations, we used one pass of Rosetta design⁵ to generate repeated sequences.

4 Packing filters – Low threshold

After completing sequence design the models were filtered out if the helices were either too far apart, creating cavities in the core (poor Rosetta holes^{S5} score, > 1.75), or too close together with an alanine-rich unspecific core packing (%alanine residues > 25%). Of the 66,776 structures that passed centroid 11,243 pass this filter.

5 Structure Profile

The structure profile biases the sequence composition towards the sequences in native proteins with similar local structure. To construct the structural profile, the sequences from the closest 100 9-residue fragments within 0.5Å RMSD to the designed structure were used. The code to construct the structural profile is included with Rosetta as generate_struct_profile.rb in tools/fragment_tools/pdb2vall. The structure profile was used in the same way as the sequence profile described by Parmeggiani *et al.*^{S7}

6 Sequence design – multipass

Starting from the filtered backbone conformations, we used Rosetta design to generate repeated sequences while minimizing the overall energy^{S4,S8}, increasing core packing as measured by Rosetta holes^{S6} and improving the psipred secondary structure prediction^{S9}. After the first round of sequence refinement the N and C terminal repeats (capping repeats) display exposed hydrophobic residues. The sequence design procedure was rerun for these repeats without a symmetric sequence to introduce polar amino acids.

7 Packing filters –High threshold

After completing sequence design the models were filtered out for poor packing. (holes score, < 0.5). After this stage we obtained 1980 structures.

8 Exploration of the energy landscape

The designs were validated using Rosetta *ab initio* structure prediction using Rosetta@Home^{S10,S11}. In Rosetta *ab initio* prediction the energy landscape is explored using independent simulations starting from an extended structure. The distribution of the simulation results is expressed in terms of energy and distance from the target fold as root mean square deviation (RMSD). A successful design produces a distribution in the shape of a funnel with the minimum corresponding to low energy and low RMSD models and no alternative minima.

For each structure, seven family members were made from the same topology, some with increased hydrogen bond potential. Proteins where multiple family members had successful simulations were selected. The member of the family with the tightest folding funnel was chosen by visual inspection and the corresponding gene was ordered for experimental testing. Extended data Fig. 3 illustrates the folding funnel and sequence diversity for one topology.

For the database we have 761 structures that have at least one family member < 3.0 RMSD from the design.

9 Add disulphides

Additional, versions with stabilizing inter-repeat disulphide bonds were also generated. Potential disulphides were scored using RosettaRemodel^{S12} and if the disulphide score was < 0 they were considered.

Time estimates

Backbone design: on a single core of a Xeon E5-2650 took 104.5 seconds to build a structure with a 19H-2L-20H-3L topology, the median topology in the database. With an average design time of 104.5 seconds per model, it would take 3493 compute days on a single core to generate the 2.8 million structures.

Sequence design – multipass: the multipass design of sequence and capping residues takes 2.1 hours for a model with 17 length helices and 3 length loops on a single core of a Xeon E5-2650.

Exploration of the energy landscape: on a single core of a Xeon E7-2850 @ 2.00 GHZ a model with 17 residues helices and 3 residues loops is produced in 19.7 minutes. Where the computation was run on Rosetta@Home, the average was 26.7 minutes. With 7 sequences per family and a minimum of 1000 models to suitably explore the landscape it would take 130 compute days per structure.

Geometrical parameters of Designed Helical Repeat proteins

- 1) Global parameters
- 2) Extracting parameters from naturally occurring repeats
- 3) Local parameters

5 1) Global parameters

Class 3 repeat proteins, as described by Kajava A.^{S13}, form solenoid structures that can be described in term of global helical parameters that relate the position of one repeat to the next one: radius (r), twist or angle between adjacent repeats around the helical axis (twist, ω) and translation between adjacent repeats along the helical axis (z).

10 Parameters for Designed Helical Repeat proteins (DHRs) and crystal structures, together with the C α RMSD values were measured on the two central repeats using the RepeatParameter filter available in Rosetta.

Radius and twist are inversely correlated and their distribution of whole set describes a hyperbolic shape, which can be represented as two symmetric ones, when considering the handedness of the superhelix in the ω value. Handedness refers to the superhelix described by the center of mass of the repeats. z is broadly distributed, with maximum values around 16 Å.

2) Extracting parameters from naturally occurring repeats

A set of alpha-helical solenoid proteins were curated from the repeatsDB (category III.3.)^{S14} to remove both proteins that had above 90% sequence identity^{S15,S16} and previously designed repeat proteins. After curation, 258 proteins remained out of 923. We then automatically extracted repeat units, which consisted of 3 subsequent repeats, that differed by less than 3 residues in length and had a high degree of structural similarity as measured by having a TM score^{S17} of greater than 0.75. The requirement of high structural similarity cut down the number of repeat proteins to 81. Repeat units were identified by the method described by RAPHAEL^{S18} implemented in Rosetta and improved. This method measures the distance from residues in the protein to random points placed around the protein. Equally spaced inflection points, where a residue was furthest or closest to these random points indicated the start of a repeat.

30 We found that inflection points occurred at random in repeat protein loops. To ensure each repeat was cut at the same location, the first residue in each repeat was chosen to be the loop-helix transition closest to the transition point. The code for this is available as extractNativeRepeats in Rosetta after git branch c876538. After locating repeats we assigned the class name of each repeat based on the PDB assignment in the Pfam database^{S19}. The Rise/Omega/Twist parameters were calculated by superimposing the first repeat-unit onto the second using TM-align^{S17} then calling the parameter calculators and averaging the values

within the same protein. This approach does not provide an extensive coverage of all the possible curvatures for each family but an indication of the protein average values.

3) Local parameters

Local parameters describe the helix-helix interactions and, due to the repeating structures, only two interactions are needed to capture the local geometry: helix1.1-helix1.2 within a repeat and helix1.1-helix2.1 between first and second repeat. Angle between helices and distance between helix centers of mass were used as parameters, extracted with a modified version of the publicly available script that can be found at the web site pymolwiki. Secondary structure definition were assigned using DSSP^{S20}. For the two central repeats, all atoms RMSDs between crystal structures and design are reported. Repeat handedness, as defined by Kobe and Kajava^{S21}, indicates the rotation of the main chain going from the N- to the C-terminal around the axis connecting the repeat centers of mass.

Structure and sequence comparison

Structural comparison of experimentally validated designs with representative repeat proteins from repeatDB^{S14} revealed that DHRs cluster in different families than the existing repeat proteins. Additionally, designs are equally distributed between right-handed and left-handed architecture, as referred to the repeat handedness (see local parameters above), in contrast to known alpha helical repeat proteins, which are mostly right-handed. This result indicates that the handedness observed is not an intrinsic limitation of repeat proteins structures but the result of a bias during evolution.

Structure determination remarks

Due to the presence of 6 cysteine residues in the native protein, the DHR5 structure was solved by sulfur single wavelength anomalous dispersion (S-SAD) using a dataset collected at 7235 eV. A search for 6 individual sulfur atoms in SHELXD gave many clear solutions that led to near complete autobuilding of a poly-alanine backbone in SHELXE, which was further elaborated using the Autobuild module of Phenix. Ultimately, the final model for DHR5 was in good agreement with the design target structure, despite our initial difficulties in phasing by molecular replacement. While the SAD data set was limited to 1.85 Å, the final model was refined against the original data set (1.25Å). Both data sets were deposited in the Protein Data Bank.

The asymmetric unit for DHR8 was found to contain 4 copies of DHR8. Although the overall structure of the 4 copies is similar, the electron density for the N-terminal helix from

two of these monomers is weak, suggesting that these helices are partially disordered in the crystal. Indeed, crystal packing of these helices in the designed conformation would have led to significant steric overlap with one another. As the corresponding helices in the remaining two DHR8 monomers were well-ordered and essentially as designed, these fully ordered

5 models were used for further analysis.

The dataset collected for DHR14 had a large non-origin Patterson peak at fractional coordinates (0.000, 0.217, 0.000), suggesting the presence of translational NCS. However, consideration of the apparent space group, unit cell parameters, and plausible solvent content strongly indicated the presence of a single copy of DHR14 in the asymmetric unit. Given the
10 relatively low pitch of this helical design and the translational pseudosymmetry between the N- and C-terminal halves of the protein, we suspected that intramolecular pseudotranslational NCS might account for the observed Patterson peak. Ultimately, a molecular replacement solution was obtained using 4 of the 8 designed helices of DHR14, and this was sufficient to bootstrap autobuilding of the remaining backbone using SHELXE. In the final model, the
15 helical axis of DHR14 is closely aligned with the crystallographic b axis, and pseudotranslational NCS between the N- and C-terminal repeats with a translation of ~ 21 Å is in good agreement with the observed fractional Patterson peak at ~ 0.22 along b.

Small Angle X-ray Scattering (SAXS) analysis

20 Guinier and $P(r)$ analysis were done using using ATSAS²⁶. The Porod exponent was determined from a linear regression analysis ($\ln$ vs q) of the top of the first peak in the Porod-Debye plot ($q^4 \cdot I(q)$ vs q^4) of the scattering data, implemented in SCATTER, available at beamline 12.3.1^{S27,S28}. The molecular mass in solution was calculated using SCATTER^{S29}.

25 25% of the designs had molecular weights in solution that were significantly greater than the predicted molecular weight (1.2-4 fold), suggesting that these designs formed multimeric assemblies or a small portion of aggregates^{S29}. All 55 designs had Porod exponents (P_E) greater than 2.9, indicating significant levels of folded protein; 67% of the designs had a P_E of 3.4-4, indicating a well-folded core^{S28}. Of the 15 proteins that crystallized, the majority (66%) had P_E of 3.9-4, consistent with more well-packed proteins
30 being easier to crystallize.

Radius of gyration (R_g) and maximum of distance distribution (d_{max}) were calculated from real space distance distribution $P(r)$. Among the models confirmed by crystallography, DHR 49 and 76 formed dimers in solution. The experimental data were fit using models based on the dimer configuration observed in the crystal structure. DHR 5

tendency to aggregation (see SEC in supporting_experimental_data.pdf) affected the SAXS profile resulting in a high Molecular weight and Vr above our acceptance threshold.

If molecular mass and Rg of models were within a 25% error from experimental data and Vr was below 2.5, the models were considered able to recapture the SAXS data. Dmax

5 errors are generally within 25%.

43 designs satisfied our requirements: DHR 1 2 3 4 7 8 9 10 14 15 18 20 21 23 24 26
27 31 32 36 39 46 47 49 52 53 54 55 57 58 59 62 64 68 70 71 72 76 77 78 79 80 81 82.

Table 3 | Protein Sequences (including optional His-tags at C-terminus)

name	sequence
DHR 1	MGCDQVAKDASSTIREVTEKNPNYSEKVADVAAKIVKKIIEGNPNGCDCVAKAASSIIRAVIEKNPNYSEV VADVAAAIIVKAIIEGNPNGCDCVAKAASSIIRAVIEKNPNYSEVADVAAAIIVKAIIEGNPNGRDCVRKAA SSIIIRAVQEKPNYSEVVEDVKRAIEKAIKEGNPNGWLEHHHHHH (SEQ ID NO: 498)
DHR 2	MSDADBAEAKAENKARNRNDDEAAKAVKLKEAIERAKKRNESDAVEAAKEAAKALNKALNRNDDEAA KAVALIAEAIIRALKRNESDAVEAAKEAAKALNKALNRNDDEAAKAVALIAEAIIRALKRNESDAVEKAKE AAKALNKALNRNDDEQAKHVAKQAENIIRALKRNESWLEHHHHHH (SEQ ID NO: 499)
DHR 3	MSEDTVRKIAQKCEAIRESDNCEBAARKCAKTISEATRESNSSELAVRIIAQVCSEATRESNDCECAAR ICAKIISEATRESNSSELAVRIIAQVCSEATRESNDCECAARICAKIISEATRESNSSELAKRIIKQVCSE AKRESNDTECAKRICKTKKSEAKRESNSWLEHHHHHH (SEQ ID NO: 500)
DHR 4	MSYEDECEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSSYEIVICECVARIVAEIVEAL KRSGTSEDEIAEIVARVISEVIRTLKESGSSYEIVICECVARTVAEIVEALKRSGTSEDEIAEIVARVISEV IRTLKESGSSYEIVKECVQRIVEIVEIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSSWLEHHHHHH (SEQ ID NO: 501)
DHR 5	MSSEKEELRERLVKICVENAKRKGDDTEEAAREAFELVREAAERAGIDSSEVLELAIRLIKECVENAQ REGYDISEACRAAAEAFKRVAEAAKRAGITSSSEVLELAIRLIKECVENAQREGYDISEACRAAAEAFKRVA EAAKRAGITSSSETLKRAEEIRKRVEEAQRREGNDISEACRQAAEEFRKKAELKRRGDGWLEHHHHHH (SEQ ID NO: 502)
DHR 6	MSEEKEEALKKVREAAKLGSSDEEARKCFEAREWAERTGSSAYEAAEALFKVLEAAYKLGSSAEEACEC FNQAAEWAERTGSGAYEAAEALFKVLEAAYKLGSSAEEACECFNQAAEWAERTGSGAYEAAERLFEELERA YEEGSSAEEACEEFNKKEEFAHRKGGKLEHHHHHH (SEQ ID NO: 503)
DHR 7	MSTKEDARSTCEKAARKAAESNDDEEVAKQAQKDCLEVAKQAGMPTKEAARSFCEAAARAAESNDDEEVAKI AAKACLEVAKQAGMPTKEAARSFCEAAARAAESNDDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAKR AAKESNDDEEVEKIAKKACKEVAKQAGMPWLEHHHHHH (SEQ ID NO: 504)
DHR 8	MSDEMCKVMEALKKAVELAKNNNDDEVAREIERAAKEIVEALRENNSDMAKVMLALAKAVLLAAKNNNDE VAREIARAAAEIVEALRENNSDMAKVMLALAKAVLLAAKNNNDDEVAREIARAAAEIVEALRENNSDMAK KMLELAKRVLDAAKNNNDETAREIARQAEEVEADRENNSWLEHHHHHH (SEQ ID NO: 505)
DHR 9	MSYEDEABEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSSYEIVIAEIVARIVAEIVEAL KRSGTSEDEIAEIVARVISEVIRTLKESGSSYEIVIAEIVARIVAEIVEALKRSGTSEDEIAEIVARVISEV IRTLKESGSSYEIVKEIVQRIVEIVEIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSSWLEHHHHHH (SEQ ID NO: 506)
DHR 10	MSSEKEELRERLVKIVVENAKRKGDDTEEAAREAFELVREAAERAGIDSSEVLELAIRLIKEVVENAQ REGYDISEAARAAAEAFKRVAEAAKRAGITSSSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVA EAAKRAGITSSSETLKRAEEIRKRVEEAQRREGNDISEAARQAEEFRKKAELKRRGDGWLEHHHHHH (SEQ ID NO: 507)
DHR 11	MSDADBAEAKAENKARNRNDDEAAKAVKLCKEAIERAKKRNESDAVEAAKEAAKALNKALNRNDDEAA KAVALCCEAIIRALKRNESDAVEAAKEAAKALNKALNRNDDEAAKAVALCCEAIIRALKRNESDAVEKAKE AAKALNKALNRNDDEQAKHVAKQCEENIIRALKRNESWLEHHHHHH (SEQ ID NO: 508)
DHR 12	MDDEEQCREIAEKAKQTYTDDDEIARIIEAAKQTTTDDDEICRCIAEAAKQTYTDDDEIARIIEAARQT TTDDDEICRCIAEAAKQTYTDDDEIARIIEAARQTTTDDDEIERCIEAAKQTYTDDDEIERIEKEYARRQ TTTDDGWLEHHHHHH (SEQ ID NO: 509)
DHR 13	MNAEDKAREVLKELKDEGSPPEEAARQVLKDLNREGSNAEDAAARAVLKALKDEGSPPEEAARAVLKALNRE GSNAEDAAARAVLKALKDEGSPPEEAARAVLKALNREGSNEEDASRAVLKALKDEGSPPEEAARAVEKALNR EGSNGWLEHHHHHH (SEQ ID NO: 510)
DHR 14	MSSEEVNRRVKQLAEKAKEATDKKEEVIEIVKELAEKQSTDSSELVNEIVKQLAEVAKKEATDKELVIYIVK ILAEKQSTDSSELVNEIVKQLAEVAKKEATDKELVIYIVKILAEKQSTDSSELVNEIVKQLEEVAKKEATD KELVEHIEKILELKKQSTDGWLEHHHHHH (SEQ ID NO: 511)
DHR 15	MNDERQKQREEVRKLAEELASKATDEELIKEIKKCAQLAEELASRSTNDELIKQILEVAKLAFELASKATD EELIKEIKKCCQLAFELASRSTNDELIKQILEVAKLAFELASKATDEELIKEIKKCCQLAFELASRSTNDE EIKQILETAKEAFERASKATDEEELIKEIKKCCQEKFEKSRSTNGWLEHHHHHH (SEQ ID NO: 512)

SUBSTITUTE SHEET (RULE 26)

36	KLAEKLRKSAGSSDILITAKAFVLEVRKGRNPFEAVKFAIKLAEKLRKSAGSSEQLEKATKVLVEEVKKG RNPKEAVEEAIKQAKEDRKRSNSGWLEHHHHHH (SEQ ID NO: 533)
DHR 37	MSSTERRAAQSVKKYLQQQKDPDQQAQKKAQEVKENIEKEANSSSVIRAAAACVFFYLLEQGYDPDQALKKAQ EVARNIENEANSSSVIRAAAACVFFYLLEQGYDPDQALKKAQEVARNIENEANSSSVIRAAAACVFFYLLEQGYDPDQALKKAQ QDPDKALBEARKRAQKTEKKTTSWLEHHHHHH (SEQ ID NO: 534)
DHR 38	MSSTERRAAQSVKKYLQQQKDPDQQAQKKAQEVKENIEKEANSSSVIRAAAACVFFYLLEQGYDPDQALKKAQ EVARNIENEANSSSVIRAAAACVFFYLLEQGYDPDQALKKAQEVARNIENEANSSSVIRAAAACVFFYLLEQGYDPDQALKKAQ QDCDKALBEARKRAQKTEKKTTSWLEHHHHHH (SEQ ID NO: 535)
DHR 39	MSDLQEVADRIVEQLKREGSPSEEARKEARRLIEEIKQSAGGDSSELIEVAVRIVKELEEQGRSPSEAAKEA VELIERIRRAAGGDSSELIEVAVRIVKELEEQGRSPSEAAKEAVELIERIRRAAGGDSSELIEVAVRIVKELEEQGRSPSEAAKEA ERGRSPSEAAARRAVEEIQRSVEEDGGNGWLEHHHHHH (SEQ ID NO: 536)
DHR 40	MSESDVAKRISKEAKKEGRSEEEVKELVERFREAIEKLKEQGDSEAIRVAVETIADEALREGLSPEEVVEL VERFVQAIQKLQENGESAIRVAVETIADEALREGLSPEEVVELVERFVQAIQKLQENGESDEIQKAVETAQ EQLEEGRSPEKVVETVEEQVKEVEEKQKQEGWLEHHHHHH (SEQ ID NO: 537)
DHR 41	MSDIEKAKRIADRAIDVVRKAAKEGGSPEKIREALQQAQKCAEKLIRLVKEAQSNGSSDVREARVALEA VRVVVRAAEKGGSPSEEVVEAVCRVRCRAEKLIRLVKRAEESNSSDVREARVALEAVRVVVRAAEKGGSPSEEVVEAVCRVRCRAEKLIRLVKRAEESNS PEEVVBAVCRVRCRAEKLIRLVKRAEESNSBNVRESARRALEKVLKTVQQAEBRGKSPSEEVVEQVCRSVRK AEEQIRETOERERSTSGWLEHHHHHH (SEQ ID NO: 538)
DHR 42	MSDAEBVKKQAEBIANRAYKTAQKQGESDSRAKKAELVRKAAEKLARLIBRAQKEGDSDALEVARQALEI ARRAFETAKKQGHSAATEAAKAFVDVVEAAISLAELIISAKRQGDSDALEVARQALEIARRAFETAKKQGHSA ATEAAKAFVDVVEAAISLAELIISAKRQGDSDALEVARQALEIARRAFETAKKQGHSAATEAAKAFVDVVEAAISLAELIISAKRQGDSDALEVARQALEI EIKKAQEQIKRERKGDGWLLEHHHHHH (SEQ ID NO: 539)
DHR 43	MSKEBELIEKARRVAKEAIEEAKRQKQDPSEAKKAELIKAVEAVKEAKRLKEEGNSSELAEILSEAIQV AVEAVEAVRQKDPFKAAEAAAEILIRAVVEAVKEAERLKREGNSSELAEILSEAIQVAVEAVEAVRQKDPFKAAEAAAEILIRAVVEAVKEAERLKREGNS PFKAAEAAAEILIRAVVEAVKEAERLKREGNSSELAKKINDTIREAVREVQQAQVEDGKDPFEAAEAAEKIRE SVERVREBEEKRRNGWLEHHHHHH (SEQ ID NO: 540)
DHR 44	MSNEQEKKDLKKAEEAAKSPDPELIREAIERABESGSKNAKEIILRAEEAAKSPDPELIRLATEAAERSG SNKAKETILRAEEAAKSPDPELIRLATEAAERSGSEKAKETIKRAEEAAKSPDPDELQKLAKAEARERLGG WLEHHHHHH (SEQ ID NO: 541)
DHR 45	MSSEEBELEKDAFEASESGADPEWLREIVDLARESGDSEVIELAKRALEAAKSGADPEWLLRIVRQAEEESG SSEVIELAKRALEAAKSGADPEWLLRIVRQAEEESGSEVIELAKRALEAAKSGADPEWLLRIVRQAEEESGSEVIELAKRALEAAKSGADPEWLLRIVRQAEEESG WLEHHHHHH (SEQ ID NO: 542)
DHR 46	MSTKEEKERIERIEKEVRSPDPENTREAVRKAEELLRENPSAEELLRRATEAAVRAPDPPEATREAVRAA EELLRENPSAEELLRRATEAAVRAPDPPEATREAVRAAEELLRENPSAEELLRRATEAAVRAPDPPEATREAVRAA REAKRAEELRKEDPGWLEHHHHHH (SEQ ID NO: 543)
DHR 47	MSTKEEKERIERIEKEVRSPDCENIREAVRKAEELLRENPSAEELLRRATEAAVRAPDPPEATREAVRAA EELLRENPSAEELLRRATEAAVRAPDPPEATREAVRAAEELLRENPSAEELLRRATEAAVRAPDPPEATREAVRAA REAKRAEELRKEDPGWLEHHHHHH (SEQ ID NO: 544)
DHR 48	MNSREBEEAKRIVKEAKKSGFDPEEVEKALREVIRVAEETGNSEALKEALKIVEEAAKSGYDPAEAVAKALA EVIRVAEETGNSEALKEALKIVEEAAKSGYDPAEAVAKALAEVIRVAEETGNPEELKEALKRVLEAAKRGED PAQVAKELAAEIRRNQEEGGWLEHHHHHH (SEQ ID NO: 545)
DHR 49	MDSEEBEQERIRIRILKEARKSGTEESLRQAIEDVAQLAKKSQDSEVLEEAIRVILRIAKESGSEELRQAIR AVAEIAKKAQDSEVLEEAIRVILRIAKESGSEELRQAIRAVAEIAKKAQDSEVLEEAIRVILRIAKESGSEELRQAIR EAEARQAERAAEETIRRAQGWLEHHHHHH (SEQ ID NO: 546)
DHR 50	MDPEEVREVERATEEYRKNPGSDEAREQLKEAVERAEEAARSPPDEAVQVAVEAATQIYENTPGSEAAK ALEIAVRAAENAAARLPDPEAVQVAVEAATQIYENTPGSEAAKALEIAVRAAENAAARLPDPEAVRVAEEAA DQIRKNTPGSELAKRADEIKERARELLERLPGWLEHHHHHH (SEQ ID NO: 547)
DHR 51	MQSEDRKREKIRELKERKARENTGSDEARQAVKEIARIKEALEEGCCDTAKEAIQRLEDLARDYSGSDVASL AVKAIKIAETALRNGYADTAKEAIQRLEDLARDYSGSDVASLAVKAIKIAETALRNGYKETAEEAIKRL RELAEYKGSSEVAKLAEEAIERIEKVSRRERGWLLEHHHHHH (SEQ ID NO: 548)
DHR 52	MQCEDRKREKIRELKERKARENTGSDEARQAVKEIARIKEALEEGCCDTAKEAIQRLEDLARDYSGSDVASL AVKAIKIAETALRNGCCDTAKEAIQRLEDLARDYSGSDVASLAVKAIKIAETALRNGCKETAEEAIKRL RELAEYKGSSEVAKLAEEAIERIEKVSRRERGWLLEHHHHHH (SEQ ID NO: 549)
DHR 53	MSNDEKELKELIKRAEELAKSPDPEDLKEAVRLAEVVRERPGSNLAKKALEILRAAEELAKLPDPEAL KEAVKAAEKVVREQPGSNLAKKALEILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSELAKKALEIIE RAAEELKKSPDPEAQKEAKKAEQKVEERPGWLEHHHHHH (SEQ ID NO: 550)
DHR 54	MTTEDEBRELEKVARKEAAAREGNTDEVREQLQRALEIARESGTTEAVKLAEVVARVAIEAARRGNTDA VREALEVALEIARESGTTEAVKLAEVVARVAIEAARRGNTDAVREALEVALEIARESGTTEAVKLAEVVARVAIEAARRGNTDA KRVSDBAKQGNEDAVKEAEVVRKIEEESGGWLEHHHHHH (SEQ ID NO: 551)
DHR 55	MSSVAEEIEKRAKKISKELKKEGKNPEWIEELQRAADKLVEVARRATSSDALEIAKRAVKIAEELAKQGSN PKWIAELLKAAKLVEVAARATSSDALEIAKRAVKIAEELAKQGSNPKWIAELLKAAKLVEVAARATSPK ALKQAKEAVKEAEELAKKGRNPKEIAEELKKRAKEVEKLARSTGWLEHHHHHH (SEQ ID NO: 552)
DHR 56	MSSVAEEIEKRCCKISKELKKEGKNPEWIEELQRAADKLVEVARRATSSDALEIAKRCVKIAEELAKQGSN PKWIAELLKACAKLVEVAARATSSDALEIAKRCVKIAEELAKQGSNPKWIAELLKACAKLVEVAARATSPK ALKQAKECVKEAEELAKKGRNPKEIAEELKKCAKEVEKLARSTGWLEHHHHHH (SEQ ID NO: 553)
DHR 57	MSTEELKKVLERVRELSERAKESTDPPEALKTAKETIELALKAVKEDPSTDALRAVLEAVRLASEVAKRVT DPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRLASEVAKRVTDPDKALKIAKLVIELALEAVKEDP SEEAKEAVEAKRLAEVSVKRVTDPELSEKIRQLVKLEEEAQKEDPGWLEHHHHHH (SEQ ID NO: 554)

	554)
DHR 58	MSTEELKKVLERVRELCEAKESTDPBEALKIATAKEVIELALKAVKEDPSTDALRAVLEAVRCACEVAKRVT DPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAKLVIELALEAVKEDP SEEAKEAVEEAKRCAEEVSKRVTDPELSEKTRQVLKLEBEEAQKEDPGWLEHHHHHH (SEQ ID NO: 555)
DHR 59	MKTEVEKKAKEVIKEAKELAKELDSEEAKEKVVERIKEAAEAAKRAAEQKGTEVAKLALKVLEEAIELAKEN RSEEAALKVVLEIARAALAAQAABEGKTEVAKLALKVLEEAIELAKENRSEEAALKVVLEIARAALAAQAA EEGKSDEARDALRRLEEAIEBAKENRSKESLEKVVREEAKEAEQQAEDAREGGWLEHHHHHH (SEQ ID NO: 556)
DHR 60	MTDIKKKAEETIKEAKKQGSSEDAIRLAQEAQKQGTBILVRAAEIVVRAQEQGSSEDAIRLAKEASREGTDIL VRAAEIVVRAQEQGSSEDAIRLAKEASREGTPTLVKAAEKVVRAQQKGSQDTIEKAKEESREGGWLEHHHHHH H (SEQ ID NO: 557)
DHR 61	MTDIKKKAEETIKEAKKQGSSEDAIRLAQECKKQGTDICVRAAEIVVRAQEQGSSEDAIRLAKECSREGTDIC VRAAEIVVRAQEQGSSEDAIRLAKECSREGTPTCVKAAEKVVRAQQKGSQDTIEKAKEESREGGWLEHHHHHH H (SEQ ID NO: 558)
DHR 62	MDNDEKRRKRAEKALQRAQEAEEKGDEEAVRAAQEA VRAAKESGDNVDLVRKVAEQALRTAKEAEKQGNVEV AVKAAARVAVEAAKQAGDNVDLVRKVAEQALRTAKEAEKQGNVEVAVKAAARVAVRAAKQAGDQDVLKRVSEQA ERISKEAKKQGNSEVSEEAARKVADEAKKQGTGGWLEHHHHHH (SEQ ID NO: 559)
DHR 63	MDPDEDRERLKEELKKIREALREAKEKPDPEIKKRALREVLEAIRRILKLAERAGDPLAREALKEINKVI REALETAKRVDPDEVKEALRVVLEAIRAILKLAEQAGDPLAREALKEINKVIREALETAKRVDPDEVKE EALRVVLEAIRAILKLAEQAGDPLAREALBEIDKVIDEAQEISERVVPDEEVQREAEVKEADRARKKLS EQSGGWLEHHHHHH (SEQ ID NO: 560)
DHR 64	MDPEDELKRVKLVKEAEELLRQAKGSEEDLEKALRTAEAAAREAKKVLEQAEKEGDPEVALRAVELVV RVAELLRIKAKESGSEELERALRVABEAAARLAKRVLELAEKQGDPEVALRAVELVVRAELLRIKAKESG SEELERALRVABEAAARLAKRVLELAEKQGDPEVARRAVELVXRVAEELLRIARESGSEEAKEPAERVREE ARELQERVKELREREGGWLEHHHHHH (SEQ ID NO: 561)
DHR 65	MDPEDELKRVKLVKEAEELLRQCKEKGSEECLEKALRTAEAAAREAKKVLEQAEKEGDPEVALRAVELVV RVAELLRIKAKESGSEECLEALRVABEAAARLAKRVLELAEKQGDPEVALRAVELVVRAELLRIKAKESG SEECLEALRVABEAAARLAKRVLELAEKQGDPEVARRAVELVXRVAEELLRIARESGSEECLEKPAERVREE ARELQERVKELREREGGWLEHHHHHH (SEQ ID NO: 562)
DHR 66	MTSDDDKVRAEAEERVREATERIQRALKKRDTDPDARKALEAAKLLKVVVEKAKKRGTSDAIKVAEAAARVAE AIARILEALNERDTPDARKALRAAIKLAEVVYKAAESGTSDAIKVAEAAARVAEAIARI LEALNERDTPDA RKALRAAIKLAEVVYKAAESGTTAEALKVAEKAARVAEKIARILEKLNERDTPBARKKLRAIKAEAKVYKE SEQGGWLEHHHHHH (SEQ ID NO: 563)
DHR 67	MTSEIDKLIKLRQTAKEVKREAEERKRRSTDP TVREVIERLAQLALDVABEAAARLIKATTSEVAKLVK LARTAEVIREAIERAERSTDPVIRVILELARLAEEVAKEAARLIVKATTSEVAKLVKWLARTAEVIRE AIERAERSTDPVIRVILELARLAEEVAKEAARLIVKATTSEVAKLVKWEAYRAIEEIRKAEKAERSTDP NEIKKILEEARKAEAEIERAKEIVKSTGWLEHHHHHH (SEQ ID NO: 564)
DHR 68	MTPRELLEEAKERVEETRELIDKARKLQEQGNKEEAEKVLREAREQIREVTRBLEETAKNSDTPALALRAA ELLVRLIKLLIETIAKLLQEQGNKEEAEKVLREATELIRKVTELLEKIAKNSDTPALALRAAELLVRLIKLL IETIAKLLQEQGNKEEAEKVLREATELIRKVTELLEKIAKNSDTPALAKRAEELLKLIELLKEIAKLEEE GNEDEAEKVKKEAEKELEERVRELEERIRKNSDQWLEHHHHHH (SEQ ID NO: 565)
DHR 69	MNPQEDLERAKEKVVRSVEEVLQRAKEAQREGDKKEVERLTKEAENQIRKARELLERVVRQNPDDPEVLLRV AELIVRLVEVVLLELAKLAEKNGDKQOVERLIQTAEELIREARELLERVSREIPDNPEVLLRVAEELIVRLVE VYLELAKLAEKNGDKQOVERLIQTAEELIREARELLERVSREIPDNPEVLLRVAEELIVRLVEVVLLELAKLA ERNQDRDQOVERLRLQAEELRREAELEERVRRERPDGWLEHHHHHH (SEQ ID NO: 566)
DHR 70	MSTEEKIEEARQSTKEAERSLREGNPEKAREBVRRALVRELEKLARKTGSTEVLI EAARLAIEVARVAL KVGSPETAREAVRTALELVQELERQARKTGSTEVLI EAARLAIEVARVALKVGSPETAREAVRTALELVQE LERQARKTGSTDEVLRKRAELAKEVARVAKEVVGSPETARQARETABRLREELRRNREKKGGWLEHHHHHH (SEQ ID NO: 567)
DHR 71	MDPEELBRAKESLERAREASERGBEEFRKAAEKALAKRLVEQAKKEGDPPELVLEAAKVALRVAEALAA KNGDKVEFKKAAESALEVAKRLVEVASKEGDPPELVLEAAKVALRVAEALAAKNGDKVEFKKAAESALEVAKR LVEVASKEGDPPELVLEAAKVAEEVRKLAKKQGDDEVYKARETAREVKEELKRVREEKGGWLEHHHHHH (SEQ ID NO: 568)
DHR 72	MDSTKEKARQLAEAAKETAEKVGDPELIKLAEQASQEGDSEKAKAILLAEEAARVAKEVGDPELIKLALEA ARRGDSEKAKAILLAEEAARVAKEVGDPELIKLALEAARRGDSEKAKAILLAEEAARBAKERGDPEQIKKA RELAKRCGWLEHHHHHH (SEQ ID NO: 569)
DHR 73	MDAEEBAKEAIKRAQEAELARKGNPEEAKEVAAEEAREAEVREAEAEKRGDAEVLALVAIALALVAIALA EVGNPEAEAREVAERAKEAEAEVRELAEKRGDAEVLALVAIALALVAIALAEVGNPEAEAREVAERAKEIAER VRELAEKRGDARVLKLVAKALELVABALKVGNPEAEAREVEERAREIKERVRLLEKGGWLEHHHHHH (SEQ ID NO: 570)
DHR 74	MDSEADRIKKLKQKEIKEVEQEARDSNDDEEELLKRLAEALKRAAEAVKRAQESGSEAIRITKKLVKEI TEVVRBARKSTDKETIELLIRLAELARAABAVADAAGSGDSEAIRITKKLVKEITEVVRBARKSTDKETI ELLIRLAELARAABAVADAAGSGDQBAIKRIKKLVKKIIEVVRKARKSTNKKEIEKLIRKAEKLARKAEQ TAEDAKRGGWLEHHHHHH (SEQ ID NO: 571)
DHR 75	MDSEKEKATELAERAQDVASRVEEAREEGSRELITARELRAREEASQEGDSEKAKAILLAAKAVLVAV EVYERAKRGSDRELREIARELAKELAAEQGDSSEKAKAILLAAKAVLVAVEVYERAKRGSDRELREIARE LAKEALRAAQEGDSEKAKAILLAAREVLRAVEQYERAKRGDDDERERAREAREALERAREGGWLEHHHH (SEQ ID NO: 572)

	HH (SEQ ID NO: 572)
DHR 76	MNPELEBEWIRRAKEVAKVEKVAQRAEEEGNPDLRDSAKELRRAVEEAI EEAKKQGNPELVEWVARAAKVA AEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIEEAVKQGNPELVEWVARAAKVA AEVIKVAIQAEKEGNR DLFRAALELVRAVIEAIEEAVKQGNPELVERVARLAKKAAELIKRAIRAEKEGNRDERREALERVREVI ER TEELVRQGGWLEHHHHHH (SEQ ID NO: 573)
DHR 77	MNSDEBEAREWEAEEAAKEALEQAKREGDEEDARRVAEELEKQABEARRKKDSEEA EAVYWAARAVLAAL EALEQAKREGDEEDARRVAEELLRQABEAAARKNSEEA EAVYWAARAVLAAL EALEQAKREGDEEDARRVAEE LLRQABEAAARKKNPEEARAVYEAARDVLEALQRL EEAARRGDEEEERREAEERLRQABEERARKGWLEHHHH HH (SEQ ID NO: 574)
DHR 78	MNSDEBEAREWEAEEAAKEALEQAKREGDEEDARRCAEELEKQABEARRKKDSEEA EAVYWAARAVLAAL EALEQAKREGDEEDARRCAEELLRQACEAARKKNSEEA EAVYWAARAVLAAL EALEQAKREGDEEDARRCAEE LLRQACEAARKKNPEEARAVYEAARDVLEALQRL EEAARRGDEEEERREAEERLRQACERARKGWLEHHHH HH (SEQ ID NO: 575)
DHR 79	MSSDEBEARELIERAKEAAERAQEAERTGDPVRELARELKR LAQEAEEVKKRDPSSSDVNEALKLIVEA IEAAVRALEAAERTGDPVRELARELVRLAVEAAEEVQRNPSSSDVNEALKLIVEA IEAAVRALEAAERTG DPEVRELARELVRLAVEAAEEVQRNPSSSEVNEALKKIVKAIQEAVESLREABESGDPEKREKARERVREA VERAEEVQRDPSCGWLEHHHHHH (SEQ ID NO: 576)
DHR 80	MNSEELERESEEAERRLQEARKRSEEAERERGD LKELAEALIEEARAVQELARVASERGNSEEAERASEKAQ RVLEEARKVSEEAEREGDDEV LALALIAIALAVLALAEVASSRGNSSEEAERASEKAQ RVLEEARKVSEEA EQGDDEV LALALIAIALAVLALAEVASSRGNKEEAERAYEDARRVEEEARKVKESAE EQGDSEVKRLAEEA EQLAREARRHVQETRGGWLEHHHHHH (SEQ ID NO: 577)
DHR 81	MNSEELERESEEAERRLQEARKRSEEAERERGD LKELAEALIEEARAVQELARVACERGNSEEAERASEKAQ RVLEEARKVSEEAEREGDDEV LALALIAIALAVLALAEVACCRGNSEEAERASEKAQ RVLEEARKVSEEA EQGDDEV LALALIAIALAVLALAEVACCRGNKEEAERAYEDARRVEEEARKVKESAE EQGDSEVKRLAEEA EQLAREARRHVQECRGGWLEHHHHHH (SEQ ID NO: 578)
DHR 82	MNDEEVQEAVERAEELREEAELIKKARKTGDPPELLRKALEALEEAVRAVEEAIKRNPDNDEAVETAVRLA RELKKVABELQERAKKTGDPPELLKLALRALEVAVRAVELAIKSNPDNDEAVETAVRLARELKKVABELQER AKKTGDPPELLKLALRALEVAVRAVELAIKSNPDNEEAVETAKRLAEELRKVABLLERAKETGDPQLELA KRAKEVADRARELAKKSNPNGWLEHHHHHH (SEQ ID NO: 579)
DHR 83	MNDEEVQEAVERAEELREEAELIKKARKTGDPPELLRKALEALEEAVRAVEEAIKRNPDNDECVETACRLA RELKKVABELQERAKKTGDPPELLKLALRALEVAVRAVELAIKSNPDNDECVETACRLARELKKVABELQER AKKTGDPPELLKLALRALEVAVRAVELAIKSNPDNEECVETAKRLAEELRKVABLLERAKETGDPQLELA KRAKEVADRARELAKKSNPNGWLEHHHHHH (SEQ ID NO: 580)

EXAMPLE COMPUTING ENVIRONMENT

- Figure 8 is a block diagram of an example computing network. Some or all of the above-mentioned techniques disclosed herein, such as but not limited to techniques disclosed as part of and/or being performed by software, the Rosetta software suite, RosettaDesign, Rosetta applications, and/or other herein-described computer software and computer hardware, can be part of and/or performed by a computing device. For example, Figure 8 shows protein design system 802 configured to communicate, via network 806, with client devices 804a, 804b, and 804c and protein database 808. In some embodiments, protein design system 802 and/or protein database 808 can be a computing device configured to perform some or all of the herein described methods and techniques, such as but not limited to, method 1000 and functionality described as being part of or related to Rosetta. Protein database 808 can, in some embodiments, store information related to and/or used by Rosetta.
- Network 806 may correspond to a LAN, a wide area network (WAN), a corporate intranet, the public Internet, or any other type of network configured to provide a communications

path between networked computing devices. Network 806 may also correspond to a combination of one or more LANs, WANs, corporate intranets, and/or the public Internet. Although Figure 8 only shows three client devices 804a, 804b, 804c, distributed application architectures may serve tens, hundreds, or thousands of client devices. Moreover, client

5 devices 804a, 804b, 804c (or any additional client devices) may be any sort of computing device, such as an ordinary laptop computer, desktop computer, network terminal, wireless communication device (e.g., a cell phone or smart phone), and so on. In some embodiments, client devices 804a, 804b, 804c can be dedicated to problem solving / using the Rosetta software suite. In other embodiments, client devices 804a, 804b, 804c can be used as

10 general purpose computers that are configured to perform a number of tasks and need not be dedicated to problem solving / using Rosetta. In still other embodiments, part or all of the functionality of protein design system 802 and/or protein database 808 can be incorporated in a client device, such as client device 804a, 804b, and/or 804c.

COMPUTING DEVICE ARCHITECTURE

15 Figure 9A is a block diagram of an example computing device (e.g., system). In particular, computing device 900 shown in Figure 9A can be configured to include components of and/or perform one or more functions of protein design system 802, client device 804a, 804b, 804c, network 806, and/or protein database 808 and/or carry out part or all of any herein-described methods and techniques, such as but not limited to method 1000.

20 Computing device 900 may include a user interface module 901, a network-communication interface module 902, one or more processors 903, and data storage 904, all of which may be linked together via a system bus, network, or other connection mechanism 905.

User interface module 901 can be operable to send data to and/or receive data from external user input/output devices. For example, user interface module 901 can be configured to

25 send and/or receive data to and/or from user input devices such as a keyboard, a keypad, a touch screen, a computer mouse, a track ball, a joystick, a camera, a voice recognition module, and/or other similar devices. User interface module 901 can also be configured to provide output to user display devices, such as one or more cathode ray tubes (CRT), liquid crystal displays (LCD), light emitting diodes (LEDs), displays using digital light processing

30 (DLP) technology, printers, light bulbs, and/or other similar devices, either now known or later developed. User interface module 901 can also be configured to generate audible output(s), such as a speaker, speaker jack, audio output port, audio output device, earphones, and/or other similar devices.

Network-communications interface module 902 can include one or more wireless interfaces 907 and/or one or more wireline interfaces 908 that are configurable to communicate via a network, such as network 806 shown in Figure 8. Wireless interfaces 907 can include one or more wireless transmitters, receivers, and/or transceivers, such as a
5 Bluetooth transceiver, a Zigbee transceiver, a Wi-Fi transceiver, a WiMAX transceiver, and/or other similar type of wireless transceiver configurable to communicate via a wireless network. Wireline interfaces 908 can include one or more wireline transmitters, receivers, and/or transceivers, such as an Ethernet transceiver, a Universal Serial Bus (USB) transceiver, or similar transceiver configurable to communicate via a twisted pair, one or
10 more wires, a coaxial cable, a fiber-optic link, or a similar physical connection to a wireline network.

In some embodiments, network communications interface module 902 can be configured to provide reliable, secured, and/or authenticated communications. For each communication described herein, information for ensuring reliable communications (*i.e.*,
15 guaranteed message delivery) can be provided, perhaps as part of a message header and/or footer (*e.g.*, packet/message sequencing information, encapsulation header(s) and/or footer(s), size/time information, and transmission verification information such as CRC and/or parity check values). Communications can be made secure (*e.g.*, be encoded or encrypted) and/or decrypted/decoded using one or more cryptographic protocols and/or
20 algorithms, such as, but not limited to, DES, AES, RSA, Diffie-Hellman, and/or DSA. Other cryptographic protocols and/or algorithms can be used as well or in addition to those listed herein to secure (and then decrypt/decode) communications.

Processors 903 can include one or more general purpose processors and/or one or more special purpose processors (*e.g.*, digital signal processors, application specific
25 integrated circuits, etc.). Processors 903 can be configured to execute computer-readable program instructions 906 contained in data storage 904 and/or other instructions as described herein. Data storage 904 can include one or more computer-readable storage media that can be read and/or accessed by at least one of processors 903. The one or more computer-readable storage media can include volatile and/or non-volatile storage
30 components, such as optical, magnetic, organic or other memory or disc storage, which can be integrated in whole or in part with at least one of processors 903. In some embodiments, data storage 904 can be implemented using a single physical device (*e.g.*, one optical, magnetic, organic or other memory or disc storage unit), while in other embodiments, data storage 904 can be implemented using two or more physical devices.

Data storage 904 can include computer-readable program instructions 906 and perhaps additional data. For example, in some embodiments, data storage 904 can store part or all of data utilized by a protein design system and/or a protein database; *e.g.*, protein designs system 802, protein database 808. In some embodiments, data storage 904 can additionally include storage required to perform at least part of the herein-described methods and techniques and/or at least part of the functionality of the herein-described devices and networks.

Figure 9B depicts a network 806 of computing clusters 909a, 909b, 909c arranged as a cloud-based server system in accordance with an example embodiment. Data and/or software for protein design system 802 can be stored on one or more cloud-based devices that store program logic and/or data of cloud-based applications and/or services. In some embodiments, protein design system 802 can be a single computing device residing in a single computing center. In other embodiments, protein design system 802 can include multiple computing devices in a single computing center, or even multiple computing devices located in multiple computing centers located in diverse geographic locations.

In some embodiments, data and/or software for protein design system 802 can be encoded as computer readable information stored in tangible computer readable media (or computer readable storage media) and accessible by client devices 804a, 804b, and 804c, and/or other computing devices. In some embodiments, data and/or software for protein design system 802 can be stored on a single disk drive or other tangible storage media, or can be implemented on multiple disk drives or other tangible storage media located at one or more diverse geographic locations.

Figure 9B depicts a cloud-based server system in accordance with an example embodiment. In Figure 9B, the functions of protein design system 802 can be distributed among three computing clusters 909a, 909b, and 909c. Computing cluster 909a can include one or more computing devices 900a, cluster storage arrays 910a, and cluster routers 911a connected by a local cluster network 912a. Similarly, computing cluster 909b can include one or more computing devices 900b, cluster storage arrays 910b, and cluster routers 911b connected by a local cluster network 912b. Likewise, computing cluster 909c can include one or more computing devices 900c, cluster storage arrays 910c, and cluster routers 911c connected by a local cluster network 912c.

In some embodiments, each of the computing clusters 909a, 909b, and 909c can have an equal number of computing devices, an equal number of cluster storage arrays, and an equal number of cluster routers. In other embodiments, however, each computing cluster can have

different numbers of computing devices, different numbers of cluster storage arrays, and different numbers of cluster routers. The number of computing devices, cluster storage arrays, and cluster routers in each computing cluster can depend on the computing task or tasks assigned to each computing cluster.

5 In computing cluster 909a, for example, computing devices 900a can be configured to perform various computing tasks of protein design system 802. In one embodiment, the various functionalities of protein design system 802 can be distributed among one or more of computing devices 900a, 900b, and 900c. Computing devices 900b and 900c in
10 computing clusters 909b and 909c can be configured similarly to computing devices 900a in computing cluster 909a. On the other hand, in some embodiments, computing devices 900a, 900b, and 900c can be configured to perform different functions.

 In some embodiments, computing tasks and stored data associated with protein design system 802 can be distributed across computing devices 900a, 900b, and 900c based at least in part on the processing requirements of protein design system 802, the processing
15 capabilities of computing devices 900a, 900b, and 900c, the latency of the network links between the computing devices in each computing cluster and between the computing clusters themselves, and/or other factors that can contribute to the cost, speed, fault-tolerance, resiliency, efficiency, and/or other design goals of the overall system architecture.

 The cluster storage arrays 910a, 910b, and 910c of the computing clusters 909a, 909b, and 909c can be data storage arrays that include disk array controllers configured to
20 manage read and write access to groups of hard disk drives. The disk array controllers, alone or in conjunction with their respective computing devices, can also be configured to manage backup or redundant copies of the data stored in the cluster storage arrays to protect against disk drive or other cluster storage array failures and/or network failures that prevent
25 one or more computing devices from accessing one or more cluster storage arrays.

 Similar to the manner in which the functions of protein design system 802 can be distributed across computing devices 900a, 900b, and 900c of computing clusters 909a, 909b, and 909c, various active portions and/or backup portions of these components can be distributed across cluster storage arrays 910a, 910b, and 910c. For example, some cluster
30 storage arrays can be configured to store one portion of the data and/or software of protein design system 802, while other cluster storage arrays can store a separate portion of the data and/or software of protein design system 802. Additionally, some cluster storage arrays can be configured to store backup versions of data stored in other cluster storage arrays.

The cluster routers 911a, 911b, and 911c in computing clusters 909a, 909b, and 909c can include networking equipment configured to provide internal and external communications for the computing clusters. For example, the cluster routers 911a in computing cluster 909a can include one or more internet switching and routing devices
5 configured to provide (i) local area network communications between the computing devices 900a and the cluster storage arrays 901a via the local cluster network 912a, and (ii) wide area network communications between the computing cluster 909a and the computing clusters 909b and 909c via the wide area network connection 913a to network 806. Cluster routers 911b and 911c can include network equipment similar to the cluster routers 911a,
10 and cluster routers 911b and 911c can perform similar networking functions for computing clusters 909b and 909b that cluster routers 911a perform for computing cluster 909a.

In some embodiments, the configuration of the cluster routers 911a, 911b, and 911c can be based at least in part on the data communication requirements of the computing devices and cluster storage arrays, the data communications capabilities of the network
15 equipment in the cluster routers 911a, 911b, and 911c, the latency and throughput of local networks 912a, 912b, 912c, the latency, throughput, and cost of wide area network links 913a, 913b, and 913c, and/or other factors that can contribute to the cost, speed, fault-tolerance, resiliency, efficiency and/or other design goals of the moderation system architecture.

20

EXAMPLE OPERATIONS

Figure 10 is a flow chart of an example method 1000. Method 1000 can begin at block 1010, where a computing device, such as computing device 900 described in the context of at least Figure 9A, can determine a protein repeating unit, where the protein
25 repeating unit can include one or more protein helices and one or more protein loops, such as discussed above at least in the context of the “Computational protocol” section. In some embodiments, the protein repeating unit can include two protein helices and two protein loops, such as discussed above at least in the context of the “Computational protocol” section.

30 In other embodiments, determining the protein repeating unit can include: selecting one or more protein fragments, each protein fragment including a plurality of protein residues; and assembling the one or more protein fragments into at least part of the protein repeating unit, such as discussed above at least in the context of the “Computational protocol” section. In particular of these embodiments, assembling the one or more protein

fragments into at least part of the protein repeating unit can include at least one of:
assembling the one or more protein fragments into a helix of the protein repeating unit and
assembling the one or more protein fragments into a loop of the protein repeating unit, such
as discussed above at least in the context of the “Computational protocol” section. In other
5 particular of these embodiments, the one or more protein fragments can include a particular
protein fragment, where each protein residue of the plurality of protein residues for the
particular protein fragment can be associated with a protein residue position; then,
determining the protein repeating unit can further include: selecting a native protein
fragment from among a plurality of native protein fragments, where the native protein
10 fragment can include a plurality of native protein residues, and where each native protein
residue of the plurality of native protein residues for the native protein fragment can be
associated with a native protein residue position, determining whether each protein residue
position associated with the plurality of particular residue positions is within a threshold
distance of a native protein residue position associated with the plurality of native protein
15 residues; and after determining that each protein residue position associated with the
plurality of particular residue positions is within the threshold distance of a native protein
residue position associated with the plurality of native protein residues, assembling the
particular protein fragment into at least part of the protein repeating unit, such as discussed
above at least in the context of the “Computational protocol” section.

20 At block 1020, the computing device can generate a protein backbone structure that
includes at least one copy of the protein repeating unit, such as discussed above at least in
the context of the “Computational protocol” section.

In some embodiments, generating the plurality of protein sequences based on the
protein backbone structure can include generating the plurality of protein sequences based
25 on the protein backbone structure such that an overall energy of the protein backbone
structure is minimized, such as discussed above at least in the context of the
“Computational protocol” section. In other embodiments, generating the plurality of protein
sequences based on the protein backbone structure can include generating the plurality of
protein sequences based on the protein backbone structure such that a core packing of the
30 protein backbone structure is increased, such as discussed above at least in the context of
the “Computational protocol” section. In still other embodiments, generating the plurality of
protein sequences based on the protein backbone structure can include generating the
plurality of protein sequences so that one or more polar amino acids is introduced into the
protein backbone structure, such as discussed above at least in the context of the

“Computational protocol” section. In even other embodiments, generating the plurality of protein sequences based on the protein backbone structure can include generating a protein sequence with one or more inter-repeat disulphide bonds, such as discussed above at least in the context of the “Computational protocol” section.

- 5 At block 1030, the computing device can determine whether a distance between a pair of helices of the protein backbone structure is between a lower distance threshold and an upper distance threshold, such as discussed above at least in the context of the “Computational protocol” section.

- 10 At block 1040, after determining that the distance between the pair of helices of the protein backbone structure is between the lower distance threshold and the upper distance threshold, the computing device can: generate a plurality of protein sequences based on the protein backbone structure, select a particular protein sequence of the plurality of protein sequences based on an energy landscape for the particular protein sequence, where the energy landscape includes information about energy and distance from a target fold of the
15 particular protein sequence, and generate an output based on the particular protein sequence, such as discussed above at least in the context of the “Computational protocol” section. In some embodiments, generating the output based on the particular protein sequence can include generating a display that includes at least part of the particular protein sequence, such as discussed above at least in the context of the “Computational protocol” section.

- 20 In some embodiments, method 1000 can further include: generating a synthetic gene encoding the particular protein sequence; expressing a particular protein *in vivo* using the synthetic gene; and purifying the particular protein, such as discussed above at least in the context of the “EXAMPLES” and “Protein expression and characterization” sections. In particular of these embodiments, expressing the particular protein sequence *in vivo* using
25 the synthetic gene can include expressing the particular protein sequence in one or more *Escherichia coli* that include the synthetic gene, such as discussed above at least in the context of the “EXAMPLES” and “Protein expression and characterization” sections. . In other particular of these embodiments, method 1000 can further include: purifying the particular protein via affinity chromatography, such as discussed above at least in the
30 context of the “EXAMPLES” and “Protein expression and characterization” sections. In still other particular of these embodiments, method 1000 can further include: synthesizing a protein having the particular protein sequence, such as discussed above at least in the context of the “EXAMPLES” and “Protein expression and characterization” sections.

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.

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).

As used herein and unless otherwise indicated, the terms "a" and "an" are taken to mean "one", "at least one" or "one or more". Unless otherwise required by context, singular terms used herein shall include pluralities and plural terms shall include the singular. 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 or 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 this application.

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 embodiments 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.

All of the references cited herein are incorporated by reference. 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.

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.

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.

With respect to any or all of the ladder diagrams, scenarios, and flow charts in the figures and as discussed herein, each block and/or communication may represent a processing of information and/or a transmission of information in accordance with example embodiments. Alternative embodiments are included within the scope of these example embodiments. In these alternative embodiments, for example, functions described as blocks, transmissions, communications, requests, responses, and/or messages may be executed out of order from that shown or discussed, including substantially concurrent or in reverse order, depending on the functionality involved. Further, more or fewer blocks and/or functions may be used with any of the ladder diagrams, scenarios, and flow charts discussed herein, and these ladder diagrams, scenarios, and flow charts may be combined with one another, in part or in whole.

A block that represents a processing of information may correspond to circuitry that can be configured to perform the specific logical functions of a herein-described method or technique. Alternatively or additionally, a block that represents a processing of information may correspond to a module, a segment, or a portion of program code (including related
5 data). The program code may include one or more instructions executable by a processor for implementing specific logical functions or actions in the method or technique. The program code and/or related data may be stored on any type of computer readable medium such as a storage device including a disk or hard drive or other storage medium.

The computer readable medium may also include non-transitory computer readable
10 media such as computer-readable media that stores data for short periods of time like register memory, processor cache, and random access memory (RAM). The computer readable media may also include non-transitory computer readable media that stores program code and/or data for longer periods of time, such as secondary or persistent long term storage, like read only memory (ROM), optical or magnetic disks, compact-disc read
15 only memory (CD-ROM), for example. The computer readable media may also be any other volatile or non-volatile storage systems. A computer readable medium may be considered a computer readable storage medium, for example, or a tangible storage device. Moreover, a block that represents one or more information transmissions may correspond to information transmissions between software and/or hardware modules in the same physical
20 device. However, other information transmissions may be between software modules and/or hardware modules in different physical devices. Numerous modifications and variations of the present disclosure are possible in light of the above teachings.

REFERENCES

1. Kajava, A. V. Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.* **179**, 279–288 (2012).
2. Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
3. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582 (2004).
4. Varadamsetty, G., Tremmel, D., Hansen, S., Parmeggiani, F. & Plückthun, A. Designed Armadillo Repeat Proteins: Library Generation, Characterization and Selection of Peptide Binders with High Specificity. *J. Mol. Biol.* **424**, 68–87 (2012).
5. Cortajarena, A. L., Liu, T. Y., Hochstrasser, M. & Regan, L. Designed Proteins To Modulate Cellular Networks. *ACS Chem. Biol.* **5**, 545–552 (2010).
6. Kobe, B. & Kajava, A. V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).
7. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. & Plückthun, A. Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. *J. Mol. Biol.* **376**, 241–257 (2008).
8. Cortajarena, A. L. & Regan, L. Calorimetric study of a series of designed repeat proteins: Modular structure and modular folding. *Protein Sci.* **20**, 336–340 (2011).
9. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *J. Mol. Biol.* **332**, 489–503 (2003).
10. Mosavi, L. K., Minor, D. L. & Peng, Z. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci.* **99**, 16029–16034 (2002).
11. Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. Design of Stable α -Helical Arrays from an Idealized TPR Motif. *Structure* **11**, 497–508 (2003).
12. Urvoas, A. *et al.* Design, Production and Molecular Structure of a New Family of Artificial Alpha-helical Repeat Proteins (α Rep) Based on Thermostable HEAT-like Repeats. *J. Mol. Biol.* **404**, 307–327 (2010).
13. Lee, S.-C. *et al.* Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc. Natl. Acad. Sci.* **109**, 3299–3304 (2012).
14. Parmeggiani, F. *et al.* Designed Armadillo Repeat Proteins as General Peptide-Binding Scaffolds: Consensus Design and Computational Optimization of the Hydrophobic Core. *J. Mol. Biol.* **376**, 1282–1304 (2008).
15. Yadid, I. & Tawfik, D. S. Reconstruction of Functional β -Propeller Lectins via Homo-oligomeric Assembly of Shorter Fragments. *J. Mol. Biol.* **365**, 10–17 (2007).
16. Coquille, S. *et al.* An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.* **5**, (2014).
17. Rämisch, S., Weininger, U., Martinsson, J., Akke, M. & André, I. Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc. Natl. Acad. Sci.* **111**, 17875–17880 (2014).
18. Lee, J. & Blaber, M. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc. Natl. Acad. Sci.* **108**, 126–130 (2011).

19. Voet, A. R. D. *et al.* Computational design of a self-assembling symmetrical β -propeller protein. *Proc. Natl. Acad. Sci.* **111**, 15102–15107 (2014).
20. Parmeggiani, F. *et al.* A General Computational Approach for Repeat Protein Design. *J. Mol. Biol.* **427**, 563–575 (2015).
- 5 21. Tripp, K. W. & Barrick, D. Enhancing the Stability and Folding Rate of a Repeat Protein through the Addition of Consensus Repeats. *J. Mol. Biol.* **365**, 1187–1200 (2007).
22. Park, K. *et al.* Control of repeat-protein curvature by computational protein design. *Nat. Struct. Mol. Biol.* **22**, 167–174 (2015).
23. Huang, P.-S. *et al.* RosettaRemodel: A Generalized Framework for Flexible
10 Backbone Protein Design. *PLoS ONE* **6**, e24109 (2011).
24. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
25. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
- 15 26. Bradley, P., Misura, K. M. S. & Baker, D. Toward High-Resolution *de Novo* Structure Prediction for Small Proteins. *Science* **309**, 1868–1871 (2005).
27. Rambo, R. P. & Tainer, J. A. Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. *Annu. Rev. Biophys.* **42**, 415–441 (2013).
28. Hura, G. L. *et al.* Robust, high-throughput solution structural analyses by small angle
20 X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612 (2009).
29. Hura, G. L. *et al.* Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods* **10**, 453–454 (2013).
30. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- 25 31. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
32. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
33. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–
30 D301 (2012).
34. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
35. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
36. Di Domenico, T. *et al.* RepeatsDB: a database of tandem repeat protein structures. *Nucleic Acids Res.* **42**, D352–D357 (2014).
37. Kabsch, W. XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).
38. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic
40 structure determination. *Acta Crystallogr. Sect. D* **58**, 1948–1954 (2002).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
40. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* **66**, 12–21 (2010).

41. Classen, S. *et al.* Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **46**, 1–13 (2013).
42. Classen, S. *et al.* Software for the high-throughput collection of SAXS data using an enhanced Blu-Ice/DCS control system. *J. Synchrotron Radiat.* **17**, 774–781 (2010).
- 5 43. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys. J.* **105**, 962–974 (2013).
44. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid
10 computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
45. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
46. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for
15 small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).

We claim:

1. A method, comprising:
 - determining a protein repeating unit using a computing device, wherein the protein repeating unit comprises one or more protein helices and one or more protein loops;
 - 5 generating a protein backbone structure that comprises at least one copy of the protein repeating unit using the computing device;
 - determining whether a distance between a pair of helices of the protein backbone structure is between a lower distance threshold and an upper distance threshold using the computing device; and
 - 10 after determining that the distance between the pair of helices of the protein backbone structure is between the lower distance threshold and the upper distance threshold, using the computing device for:
 - generating a plurality of protein sequences based on the protein backbone structure,
 - selecting a particular protein sequence of the plurality of protein sequences based on
 - 15 an energy landscape for the particular protein sequence, wherein the energy landscape comprises information about energy and distance from a target fold of the particular protein sequence, and
 - generating an output based on the particular protein sequence.
2. The method of claim 1, wherein the protein repeating unit comprises two protein
- 20 helices and two protein loops.
3. The method of either claim 1 or claim 2, wherein determining the protein repeating unit comprises:
 - selecting one or more protein fragments, each protein fragment comprising a plurality of protein residues; and
 - 25 assembling the one or more protein fragments into at least part of the protein repeating unit.
4. The method of claim 3, wherein assembling the one or more protein fragments into at least part of the protein repeating unit comprises at least one of: assembling the one or more protein fragments into a helix of the protein repeating unit and assembling the one or more
- 30 protein fragments into a loop of the protein repeating unit.
5. The method of claim 3 or claim 4, wherein the one or more protein fragments comprise a particular protein fragment, wherein each protein residue of the plurality of

- protein residues for the particular protein fragment is associated with a protein residue position, and wherein determining the protein repeating unit further comprises:
selecting a native protein fragment from among a plurality of native protein fragments,
wherein the native protein fragment comprises a plurality of native protein residues, and
5 wherein each native protein residue of the plurality of native protein residues for the native protein fragment is associated with a native protein residue position;
determining whether each protein residue position associated with the plurality of particular residue positions is within a threshold distance of a native protein residue position associated with the plurality of native protein residues; and
10 after determining that each protein residue position associated with the plurality of particular residue positions is within the threshold distance of a native protein residue position associated with the plurality of native protein residues, assembling the particular protein fragment into at least part of the protein repeating unit.
6. The method of any one of claims 1-5, wherein generating the plurality of protein
15 sequences based on the protein backbone structure comprises:
generating the plurality of protein sequences based on the protein backbone structure such that an overall energy of the protein backbone structure is minimized.
7. The method of any one of claims 1-6, wherein generating the plurality of protein
sequences based on the protein backbone structure comprises:
20 generating the plurality of protein sequences based on the protein backbone structure such that a core packing of the protein backbone structure is increased.
8. The method of any one of claims 1-7, wherein generating the plurality of protein
sequences based on the protein backbone structure comprises generating the plurality of
protein sequences so that one or more polar amino acids is introduced into the protein
25 backbone structure.
9. The method of any one of claims 1-8, wherein generating the plurality of protein
sequences based on the protein backbone structure comprises generating a protein sequence
with one or more inter-repeat disulphide bonds.
10. The method of any one of claims 1-9, wherein generating the output based on the
30 particular protein sequence comprises generating a display that includes at least part of the particular protein sequence.
11. The method of any one of claims 1-10, further comprising:
generating a synthetic gene encoding the particular protein sequence;
expressing a particular protein *in vivo* using the synthetic gene; and

purifying the particular protein.

12. The method of claim 11, wherein expressing the particular protein sequence *in vivo* using the synthetic gene comprises expressing the particular protein sequence in one or more *Escherichia coli* that include the synthetic gene.

5 13. The method of either claim 11 or claim 12, further comprising:
purifying the particular protein via affinity chromatography.

14. The method of any one of claims 1-13, further comprising:
synthesizing a protein having the particular protein sequence.

15. A computing device, comprising:
10 one or more data processors; and
a computer-readable medium, configured to store at least computer-readable instructions that, when executed, cause the computing device to perform the method of any one of claims 1-14.

16. The computing device of claim 15, wherein the computer-readable medium comprises
15 a non-transitory computer-readable medium.

17. A computer-readable medium, configured to store at least computer-readable instructions that, when executed by one or more processors of a computing device, cause the computing device to perform the method of any one of claims 1-14.

18. The computer-readable medium of claim 17, wherein the computer-readable medium
20 comprises a non-transitory computer-readable medium.

19. A polypeptide comprising or consisting of the amino acid sequence selected from the group consisting of:

- (a) SEQ ID NO:1-[SEQ ID NO:2]_(0 or 2-19)-SEQ ID NO:3;
- (b) SEQ ID NO:7-[SEQ ID NO:8]_(0 or 2-19)-SEQ ID NO:9;
- 25 (c) SEQ ID NO:13-[SEQ ID NO:14]_(0 or 2-19)-SEQ ID NO:15;
- (d) SEQ ID NO:19-[SEQ ID NO:20]_(0 or 2-19)-SEQ ID NO:21;
- (e) SEQ ID NO:25-[SEQ ID NO:26]_(0 or 2-19)-SEQ ID NO:27;
- (f) SEQ ID NO:31-[SEQ ID NO:32]_(0 or 2-19)-SEQ ID NO:33;
- (g) SEQ ID NO:37-[SEQ ID NO:38]_(0 or 2-19)-SEQ ID NO:39;
- 30 (h) SEQ ID NO:43-[SEQ ID NO:44]_(0 or 2-19)-SEQ ID NO:45;
- (i) SEQ ID NO:49-[SEQ ID NO:50]_(0 or 2-19)-SEQ ID NO:51;
- (j) SEQ ID NO:55-[SEQ ID NO:56]_(0 or 2-19)-SEQ ID NO:57;
- (k) SEQ ID NO:61-[SEQ ID NO:62]_(0 or 2-19)-SEQ ID NO:63;
- (l) SEQ ID NO:67-[SEQ ID NO:68]_(0 or 2-19)-SEQ ID NO:69;

- (m) SEQ ID NO:73-[SEQ ID NO:74]_(0 or 2-19)-SEQ ID NO:75;
- (n) SEQ ID NO:79-[SEQ ID NO:80]_(0 or 2-19)-SEQ ID NO:81;
- (o) SEQ ID NO:85-[SEQ ID NO:86]_(0 or 2-19)-SEQ ID NO:87;
- (p) SEQ ID NO:91-[SEQ ID NO:92]_(0 or 2-19)-SEQ ID NO:93;
- 5 (q) SEQ ID NO:97-[SEQ ID NO:98]_(0 or 2-19)-SEQ ID NO:99;
- (r) SEQ ID NO:103-[SEQ ID NO:104]_(0 or 2-19)-SEQ ID NO:105;
- (s) SEQ ID NO:109-[SEQ ID NO:110]_(0 or 2-19)-SEQ ID NO:111;
- (t) SEQ ID NO:115-[SEQ ID NO:116]_(0 or 2-19)-SEQ ID NO:117;
- (u) SEQ ID NO:121-[SEQ ID NO:122]_(0 or 2-19)-SEQ ID NO:123;
- 10 (v) SEQ ID NO:127-[SEQ ID NO:128]_(0 or 2-19)-SEQ ID NO:129;
- (w) SEQ ID NO:133-[SEQ ID NO:134]_(0 or 2-19)-SEQ ID NO:135;
- (x) SEQ ID NO:139-[SEQ ID NO:140]_(0 or 2-19)-SEQ ID NO:141;
- (y) SEQ ID NO:145-[SEQ ID NO:146]_(0 or 2-19)-SEQ ID NO:147;
- (z) SEQ ID NO:151-[SEQ ID NO:152]_(0 or 2-19)-SEQ ID NO:153;
- 15 (aa) SEQ ID NO:157-[SEQ ID NO:158]_(0 or 2-19)-SEQ ID NO:159;
- (bb) SEQ ID NO:163-[SEQ ID NO:164]_(0 or 2-19)-SEQ ID NO:165;
- (cc) SEQ ID NO:169-[SEQ ID NO:170]_(0 or 2-19)-SEQ ID NO:171;
- (dd) SEQ ID NO:175-[SEQ ID NO:176]_(0 or 2-19)-SEQ ID NO:177;
- (ee) SEQ ID NO:181-[SEQ ID NO:182]_(0 or 2-19)-SEQ ID NO:183;
- 20 (ff) SEQ ID NO:187-[SEQ ID NO:188]_(0 or 2-19)-SEQ ID NO:189;
- (gg) SEQ ID NO:193-[SEQ ID NO:194]_(0 or 2-19)-SEQ ID NO:195;
- (hh) SEQ ID NO:199-[SEQ ID NO:200]_(0 or 2-19)-SEQ ID NO:201;
- (ii) SEQ ID NO:205-[SEQ ID NO:206]_(0 or 2-19)-SEQ ID NO:207;
- (jj) SEQ ID NO:211-[SEQ ID NO:212]_(0 or 2-19)-SEQ ID NO:213;
- 25 (kk) SEQ ID NO:217-[SEQ ID NO:218]_(0 or 2-19)-SEQ ID NO:219;
- (ll) SEQ ID NO:223-[SEQ ID NO:224]_(0 or 2-19)-SEQ ID NO:225;
- (mm) SEQ ID NO:229-[SEQ ID NO:230]_(0 or 2-19)-SEQ ID NO:231;
- (nn) SEQ ID NO:235-[SEQ ID NO:236]_(0 or 2-19)-SEQ ID NO:237;
- (oo) SEQ ID NO:241-[SEQ ID NO:242]_(0 or 2-19)-SEQ ID NO:243;
- 30 (pp) SEQ ID NO:247-[SEQ ID NO:248]_(0 or 2-19)-SEQ ID NO:249;
- (qq) SEQ ID NO:253-[SEQ ID NO:254]_(0 or 2-19)-SEQ ID NO:255;
- (rr) SEQ ID NO:259-[SEQ ID NO:260]_(0 or 2-19)-SEQ ID NO:261;
- (ss) SEQ ID NO:265-[SEQ ID NO:266]_(0 or 2-19)-SEQ ID NO:267;
- (tt) SEQ ID NO:271-[SEQ ID NO:272]_(0 or 2-19)-SEQ ID NO:273;

- (uu) SEQ ID NO:277-[SEQ ID NO:278]_(0 or 2-19)-SEQ ID NO:278;
 (vv) SEQ ID NO:283-[SEQ ID NO:284]_(0 or 2-19)-SEQ ID NO:285;
 (ww) SEQ ID NO:289-[SEQ ID NO:290]_(0 or 2-19)-SEQ ID NO:291;
 (xx) SEQ ID NO:295-[SEQ ID NO:296]_(0 or 2-19)-SEQ ID NO:297;
 5 (yy) SEQ ID NO:301-[SEQ ID NO:302]_(0 or 2-19)-SEQ ID NO:303;
 (zz) SEQ ID NO:307-[SEQ ID NO:308]_(0 or 2-19)-SEQ ID NO:309;
 (aaa) SEQ ID NO:313-[SEQ ID NO:314]_(0 or 2-19)-SEQ ID NO:315;
 (bbb) SEQ ID NO:319-[SEQ ID NO:320]_(0 or 2-19)-SEQ ID NO:321;
 (ccc) SEQ ID NO:325-[SEQ ID NO:326]_(0 or 2-19)-SEQ ID NO:327;
 10 (ddd) SEQ ID NO:331-[SEQ ID NO:332]_(0 or 2-19)-SEQ ID NO:333;
 (eee) SEQ ID NO:337-[SEQ ID NO:338]_(0 or 2-19)-SEQ ID NO:339;
 (fff) SEQ ID NO:343-[SEQ ID NO:344]_(0 or 2-19)-SEQ ID NO:345;
 (ggg) SEQ ID NO:349-[SEQ ID NO:350]_(0 or 2-19)-SEQ ID NO:351;
 (hhh) SEQ ID NO:355-[SEQ ID NO:356]_(0 or 2-19)-SEQ ID NO:357;
 15 (iii) SEQ ID NO:361-[SEQ ID NO:362]_(0 or 2-19)-SEQ ID NO:363;
 (jjj) SEQ ID NO:367-[SEQ ID NO:368]_(0 or 2-19)-SEQ ID NO:369;
 (kkk) SEQ ID NO:373-[SEQ ID NO:374]_(0 or 2-19)-SEQ ID NO:375;
 (lll) SEQ ID NO:379-[SEQ ID NO:380]_(0 or 2-19)-SEQ ID NO:381;
 (mmm) SEQ ID NO:385-[SEQ ID NO:386]_(0 or 2-19)-SEQ ID NO:387;
 20 (nnn) SEQ ID NO:391-[SEQ ID NO:392]_(0 or 2-19)-SEQ ID NO:393;
 (ooo) SEQ ID NO:397-[SEQ ID NO:398]_(0 or 2-19)-SEQ ID NO:399;
 (ppp) SEQ ID NO:403-[SEQ ID NO:404]_(0 or 2-19)-SEQ ID NO:405; and
 (qqq) SEQ ID NO:409-[SEQ ID NO:410]_(0 or 2-19)-SEQ ID NO:411;
 wherein the domain in brackets is an optional internal domain.

- 25 20. The polypeptide of claim 19, wherein the polypeptide comprises or consists of the amino acid sequence selected from the group consisting of:

- (A) SEQ ID NO:4-[SEQ ID NO:5]_(0 or 2-19)-SEQ ID NO:6;
 (B) SEQ ID NO:10-[SEQ ID NO:11]_(0 or 2-19)-SEQ ID NO:12;
 (C) SEQ ID NO:16-[SEQ ID NO:17]_(0 or 2-19)-SEQ ID NO:18;
 30 (D) SEQ ID NO:22-[SEQ ID NO:23]_(0 or 2-19)-SEQ ID NO:24;
 (E) SEQ ID NO:28-[SEQ ID NO:29]_(0 or 2-19)-SEQ ID NO:30;
 (F) SEQ ID NO:34-[SEQ ID NO:35]_(0 or 2-19)-SEQ ID NO:36;
 (G) SEQ ID NO:40-[SEQ ID NO:41]_(0 or 2-19)-SEQ ID NO:42;
 (H) SEQ ID NO:46-[SEQ ID NO:47]_(0 or 2-19)-SEQ ID NO:48;

- (I) SEQ ID NO:52-[SEQ ID NO:53]_(0 or 2-19)-SEQ ID NO:54;
 (J) SEQ ID NO:58-[SEQ ID NO:59]_(0 or 2-19)-SEQ ID NO:60;
 (K) SEQ ID NO:64-[SEQ ID NO:65]_(0 or 2-19)-SEQ ID NO:66;
 (L) SEQ ID NO:70-[SEQ ID NO:71]_(0 or 2-19)-SEQ ID NO:72;
 5 (M) SEQ ID NO:76-[SEQ ID NO:77]_(0 or 2-19)-SEQ ID NO:78;
 (N) SEQ ID NO:82-[SEQ ID NO:83]_(0 or 2-19)-SEQ ID NO:84;
 (O) SEQ ID NO:88-[SEQ ID NO:89]_(0 or 2-19)-SEQ ID NO:90;
 (P) SEQ ID NO:94-[SEQ ID NO:95]_(0 or 2-19)-SEQ ID NO:96;
 (Q) SEQ ID NO:100-[SEQ ID NO:101]_(0 or 2-19)-SEQ ID NO:102;
 10 (R) SEQ ID NO:106-[SEQ ID NO:107]_(0 or 2-19)-SEQ ID NO:108;
 (S) SEQ ID NO:112-[SEQ ID NO:113]_(0 or 2-19)-SEQ ID NO:114;
 (T) SEQ ID NO:118-[SEQ ID NO:119]_(0 or 2-19)-SEQ ID NO:120;
 (U) SEQ ID NO:124-[SEQ ID NO:125]_(0 or 2-19)-SEQ ID NO:126;
 (V) SEQ ID NO:130-[SEQ ID NO:131]_(0 or 2-19)-SEQ ID NO:132;
 15 (W) SEQ ID NO:136-[SEQ ID NO:137]_(0 or 2-19)-SEQ ID NO:138;
 (X) SEQ ID NO:142-[SEQ ID NO:143]_(0 or 2-19)-SEQ ID NO:144;
 (Y) SEQ ID NO:148-[SEQ ID NO:149]_(0 or 2-19)-SEQ ID NO:150;
 (Z) SEQ ID NO:154-[SEQ ID NO:155]_(0 or 2-19)-SEQ ID NO:156;
 (AA) SEQ ID NO:160-[SEQ ID NO:161]_(0 or 2-19)-SEQ ID NO:162;
 20 (BB) SEQ ID NO:166-[SEQ ID NO:167]_(0 or 2-19)-SEQ ID NO:168;
 (CC) SEQ ID NO:172-[SEQ ID NO:173]_(0 or 2-19)-SEQ ID NO:174;
 (DD) SEQ ID NO:178-[SEQ ID NO:179]_(0 or 2-19)-SEQ ID NO:180;
 (EE) SEQ ID NO:184-[SEQ ID NO:185]_(0 or 2-19)-SEQ ID NO:186;
 (FF) SEQ ID NO:190-[SEQ ID NO:191]_(0 or 2-19)-SEQ ID NO:192;
 25 (GG) SEQ ID NO:196-[SEQ ID NO:197]_(0 or 2-19)-SEQ ID NO:198;
 (HH) SEQ ID NO:202-[SEQ ID NO:203]_(0 or 2-19)-SEQ ID NO:204;
 (II) SEQ ID NO:208-[SEQ ID NO:209]_(0 or 2-19)-SEQ ID NO:210;
 (JJ) SEQ ID NO:214-[SEQ ID NO:215]_(0 or 2-19)-SEQ ID NO:216;
 (KK) SEQ ID NO:220-[SEQ ID NO:221]_(0 or 2-19)-SEQ ID NO:222;
 30 (LL) SEQ ID NO:226-[SEQ ID NO:227]_(0 or 2-19)-SEQ ID NO:228;
 (MM) SEQ ID NO:232-[SEQ ID NO:233]_(0 or 2-19)-SEQ ID NO:234;
 (NN) SEQ ID NO:238-[SEQ ID NO:239]_(0 or 2-19)-SEQ ID NO:240;
 (OO) SEQ ID NO:244-[SEQ ID NO:245]_(0 or 2-19)-SEQ ID NO:246;
 (PP) SEQ ID NO:250-[SEQ ID NO:251]_(0 or 2-19)-SEQ ID NO:252;

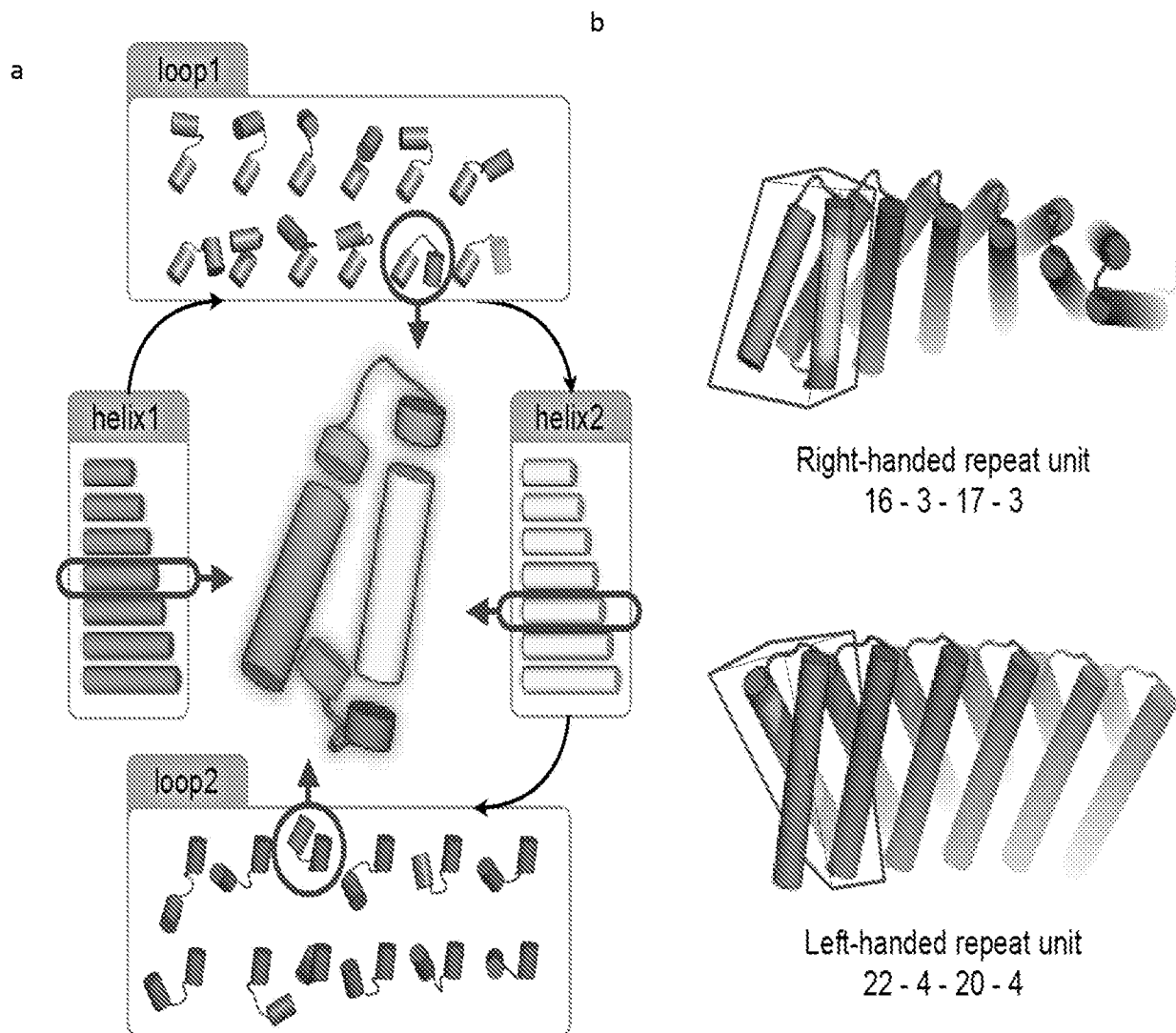
(QQ) SEQ ID NO:256-[SEQ ID NO:257]_(0 or 2-19)-SEQ ID NO:258;
 (RR) SEQ ID NO:262-[SEQ ID NO:263]_(0 or 2-19)-SEQ ID NO:264;
 (SS) SEQ ID NO:268-[SEQ ID NO:269]_(0 or 2-19)-SEQ ID NO:270;
 (TT) SEQ ID NO:274-[SEQ ID NO:275]_(0 or 2-19)-SEQ ID NO:276;
 5 (UU) SEQ ID NO:280-[SEQ ID NO:281]_(0 or 2-19)-SEQ ID NO:282;
 (VV) SEQ ID NO:286-[SEQ ID NO:287]_(0 or 2-19)-SEQ ID NO:288;
 (WW) SEQ ID NO:292-[SEQ ID NO:293]_(0 or 2-19)-SEQ ID NO:294;
 (XX) SEQ ID NO:298-[SEQ ID NO:299]_(0 or 2-19)-SEQ ID NO:300;
 (YY) SEQ ID NO:304-[SEQ ID NO:305]_(0 or 2-19)-SEQ ID NO:306;
 10 (ZZ) SEQ ID NO:310-[SEQ ID NO:311]_(0 or 2-19)-SEQ ID NO:312;
 (AAA) SEQ ID NO:316-[SEQ ID NO:317]_(0 or 2-19)-SEQ ID NO:318;
 (BBB) SEQ ID NO:322-[SEQ ID NO:323]_(0 or 2-19)-SEQ ID NO:324;
 (CCC) SEQ ID NO:328-[SEQ ID NO:329]_(0 or 2-19)-SEQ ID NO:330;
 (DDD) SEQ ID NO:334-[SEQ ID NO:335]_(0 or 2-19)-SEQ ID NO:336;
 15 (EEE) SEQ ID NO:340-[SEQ ID NO:341]_(0 or 2-19)-SEQ ID NO:342;
 (FFF) SEQ ID NO:346-[SEQ ID NO:347]_(0 or 2-19)-SEQ ID NO:348;
 (GGG) SEQ ID NO:352-[SEQ ID NO:353]_(0 or 2-19)-SEQ ID NO:354;
 (HHH) SEQ ID NO:358-[SEQ ID NO:359]_(0 or 2-19)-SEQ ID NO:360;
 (III) SEQ ID NO:364-[SEQ ID NO:365]_(0 or 2-19)-SEQ ID NO:366;
 20 (JJJ) SEQ ID NO:370-[SEQ ID NO:371]_(0 or 2-19)-SEQ ID NO:372;
 (KKK) SEQ ID NO:376-[SEQ ID NO:377]_(0 or 2-19)-SEQ ID NO:378;
 (LLL) SEQ ID NO:382-[SEQ ID NO:383]_(0 or 2-19)-SEQ ID NO:384;
 (MMM) SEQ ID NO:388-[SEQ ID NO:389]_(0 or 2-19)-SEQ ID NO:390;
 (NNN) SEQ ID NO:394-[SEQ ID NO:395]_(0 or 2-19)-SEQ ID NO:396;
 25 (OOO) SEQ ID NO:400-[SEQ ID NO:401]_(0 or 2-19)-SEQ ID NO:402;
 (PPP) SEQ ID NO:406-[SEQ ID NO:407]_(0 or 2-19)-SEQ ID NO:408; and
 (QQQ) SEQ ID NO:412-[SEQ ID NO:413]_(0 or 2-19)-SEQ ID NO:414;
 wherein the domain in brackets is an optional internal domain.

21. The polypeptide of claim 19 or 20, wherein the optional internal domain is absent.
- 30 22. The polypeptide of claim 19 or 20, wherein the optional internal domain is present in 2-19 copies.
23. The polypeptide of claim 19 or 20, wherein the optional internal domain is present in 2-3 copies.

24. A polypeptide comprising or consisting of a polypeptide having at least 50% identity over its length with the amino acid sequence selected from the group consisting of SEQ ID NO: 415-497.
25. The polypeptide of claim 24, comprising or consisting of a polypeptide having at least
5 75% identity over its length with the amino acid sequence selected from the group consisting of SEQ ID NO: 415-497.
26. The polypeptide of claim 24, comprising or consisting of a polypeptide having at least 90% identity over its length with the amino acid sequence selected from the group consisting of SEQ ID NO: 415-497.
- 10 27. The polypeptide of claim 24, comprising or consisting of the amino acid sequence selected from the group consisting of SEQ ID NO: 415-497.
28. A protein assembly comprising a plurality of polypeptides having the same amino acid sequence selected from the group listed in any of claims 19- 27.
29. A recombinant nucleic acid encoding a polypeptide of any of claims 19-28.
- 15 30. A recombinant expression vector comprising the nucleic acid of claim 29 operatively linked to a promoter.
31. A recombinant host cell comprising the recombinant expression vectors of claim 30.

20

Figure 1



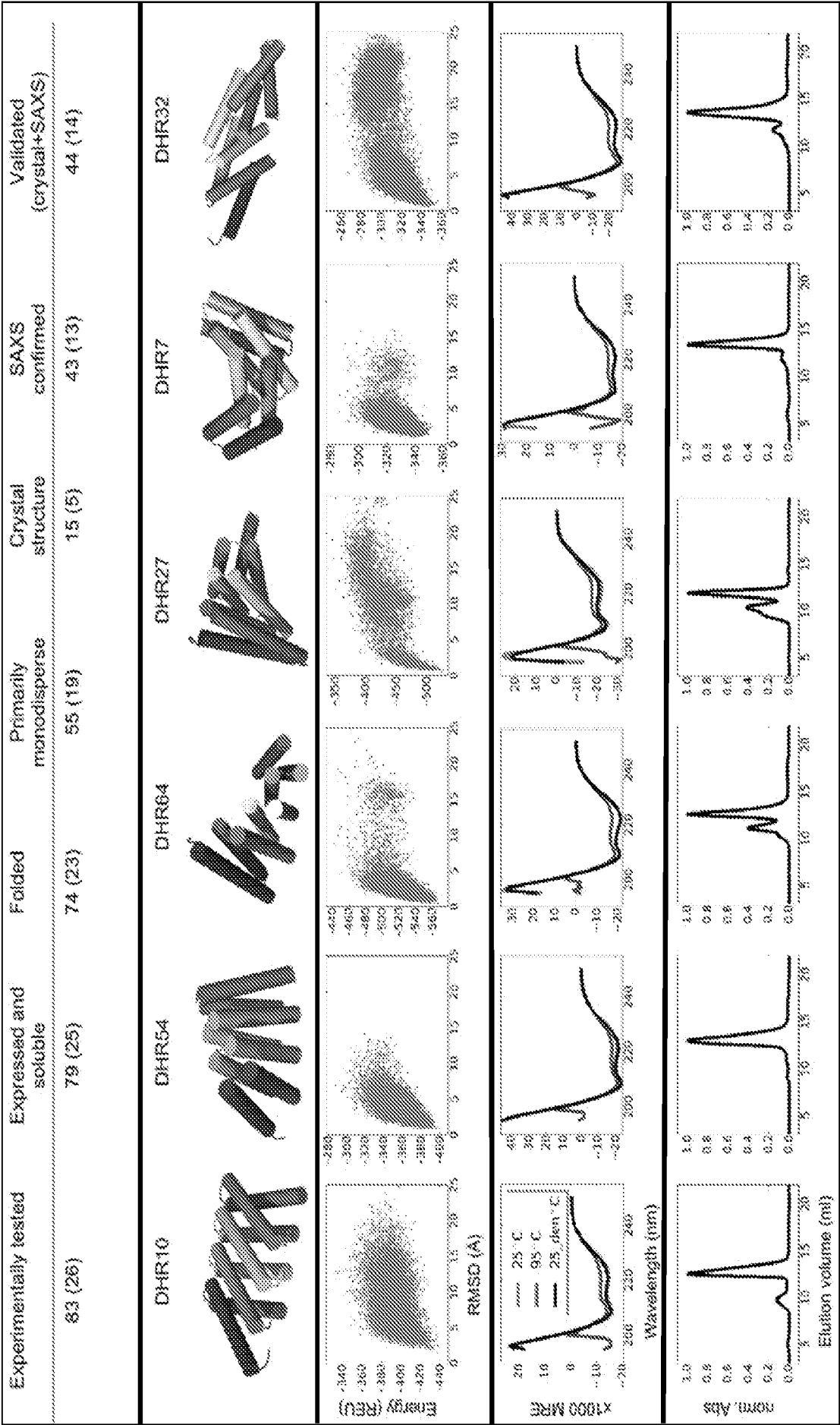
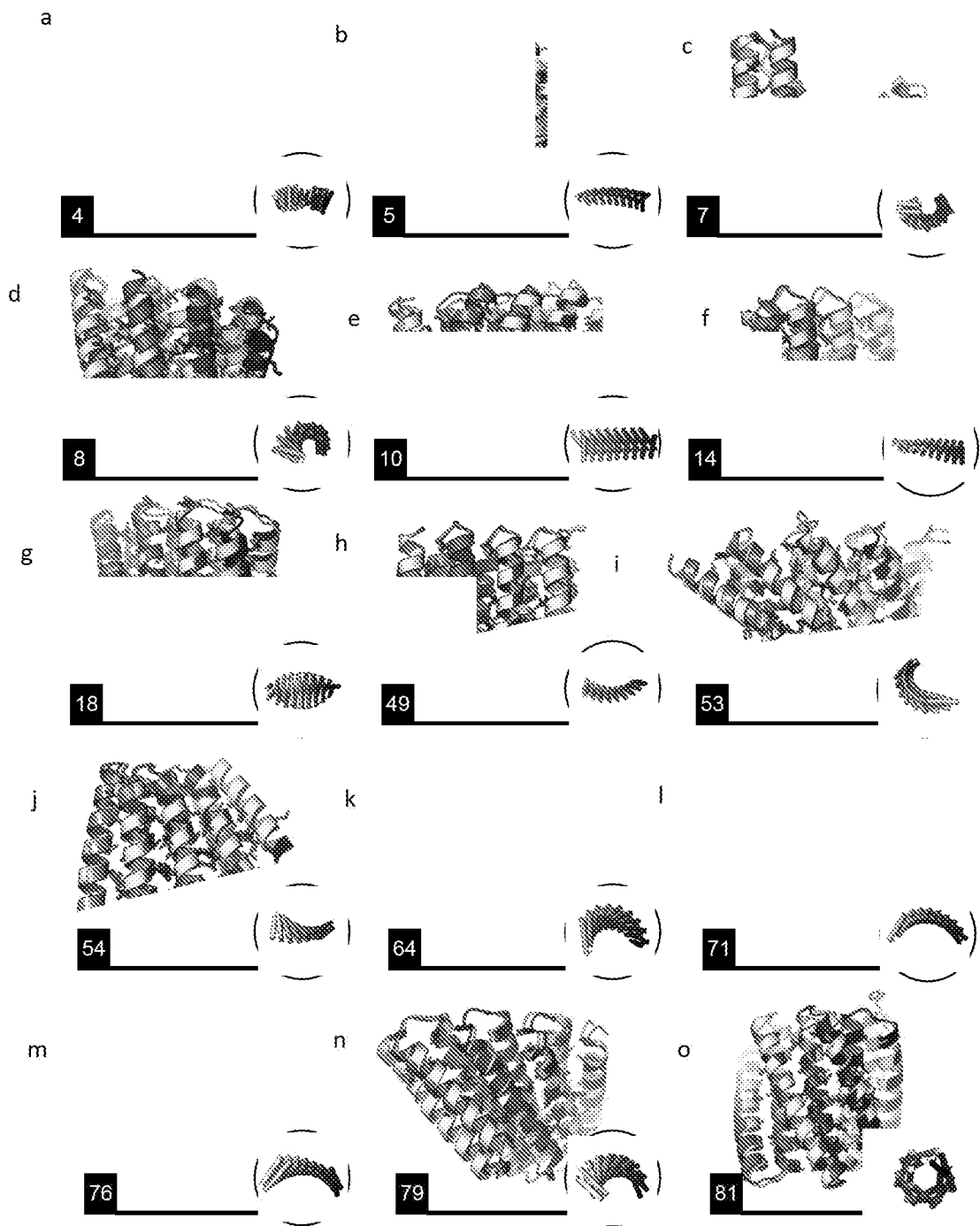


FIGURE 2

Figure 3



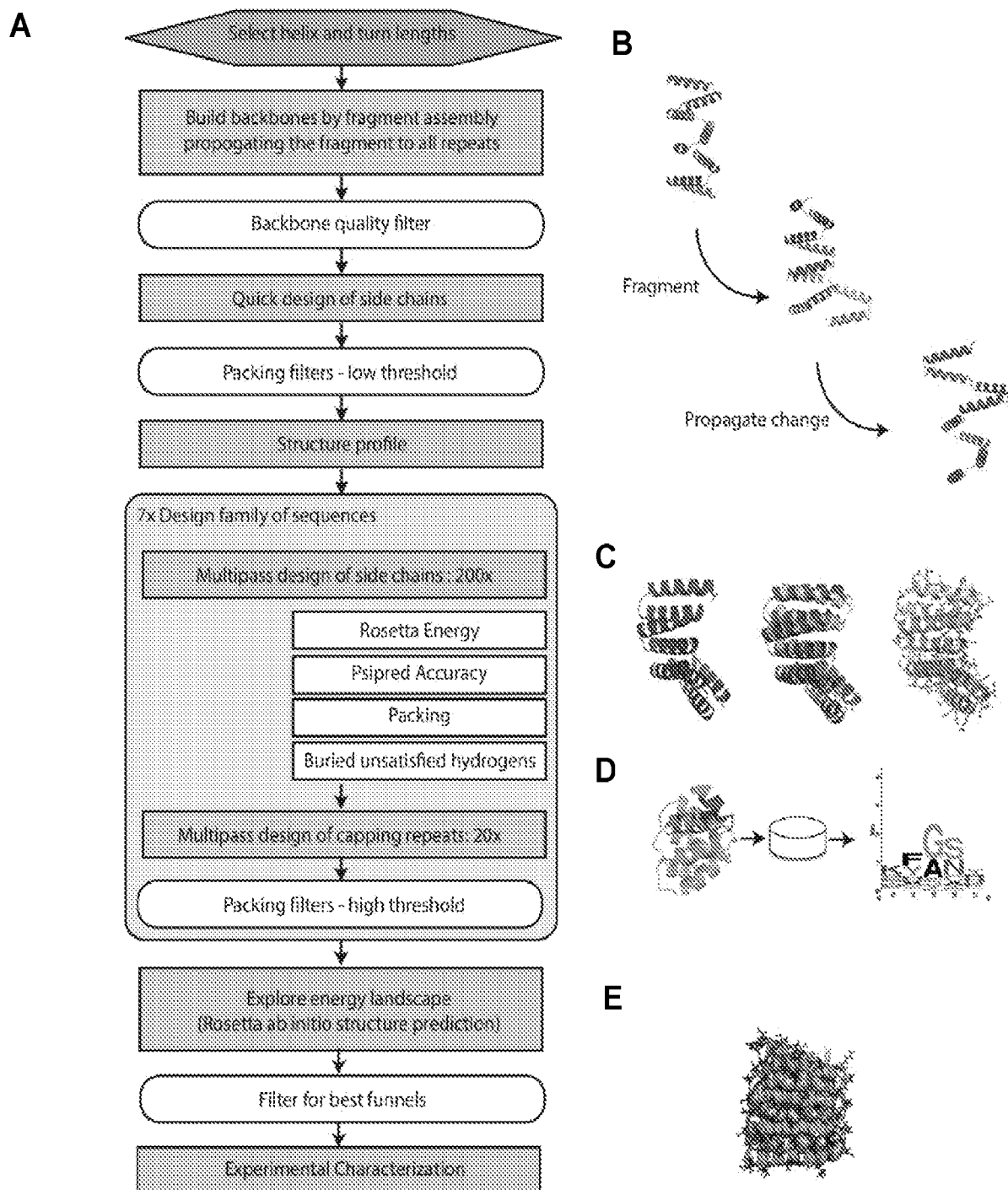


FIGURE 4

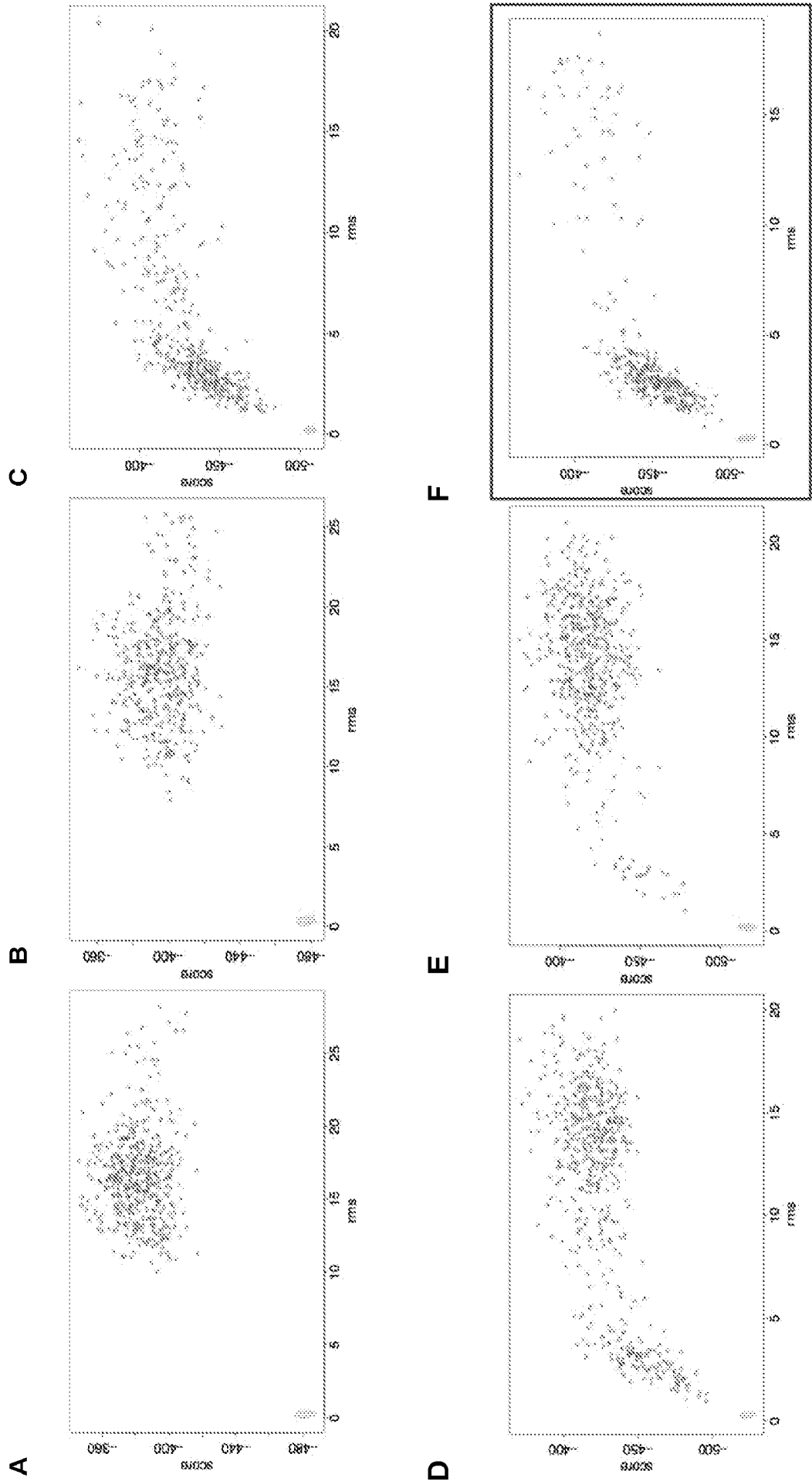
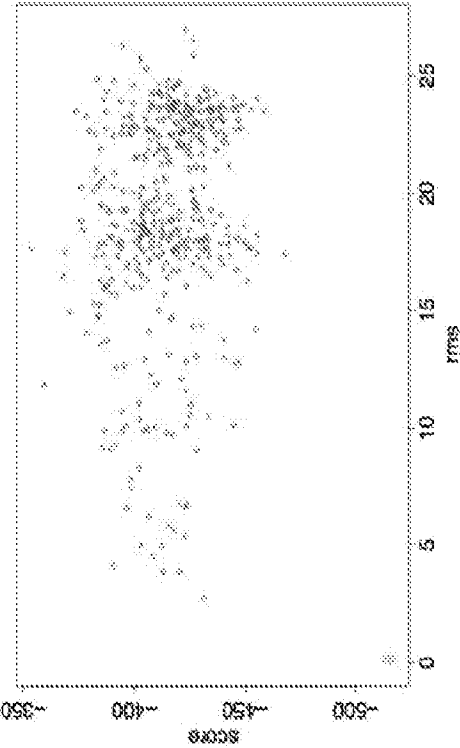


FIGURE 5A – 5F

G



H

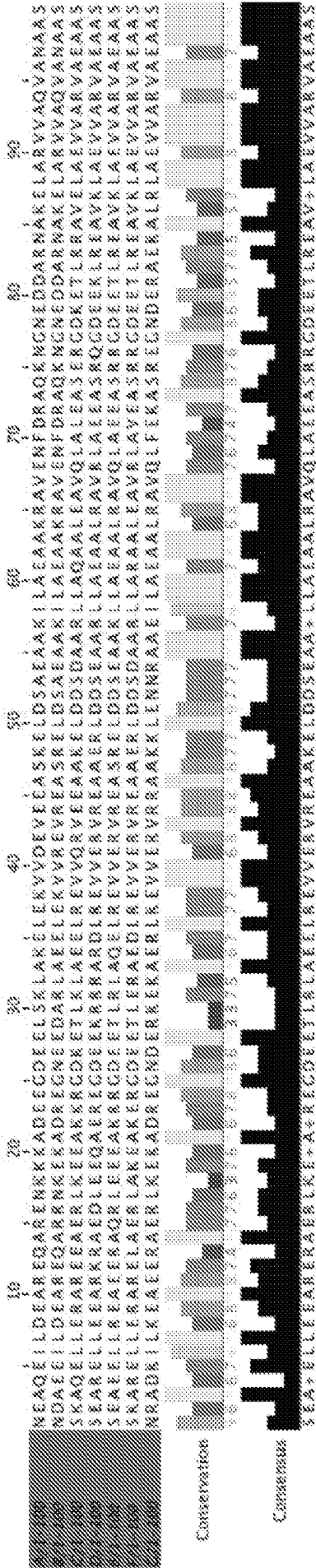


FIGURE 5G - 5H

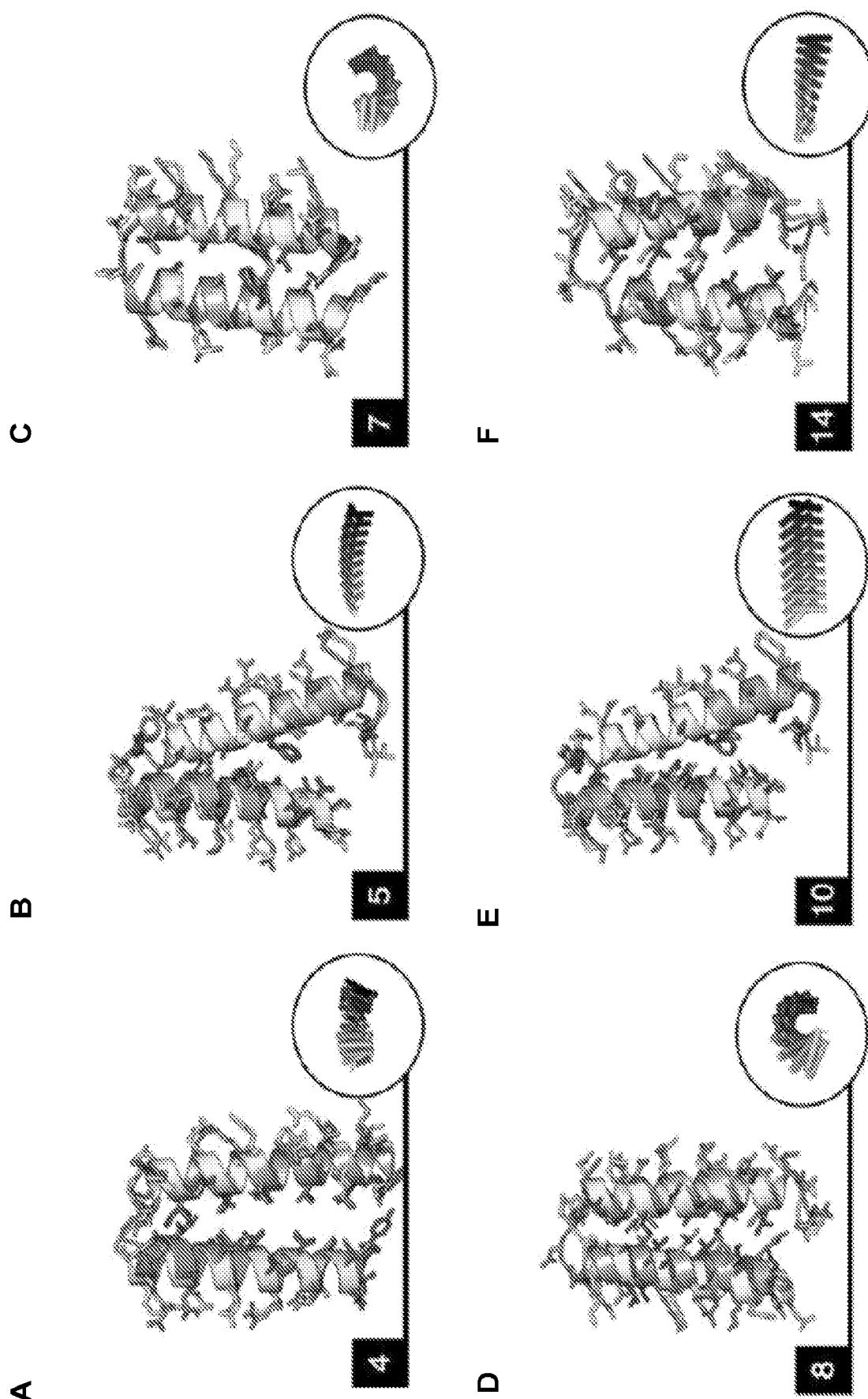


FIGURE 6A – 6F

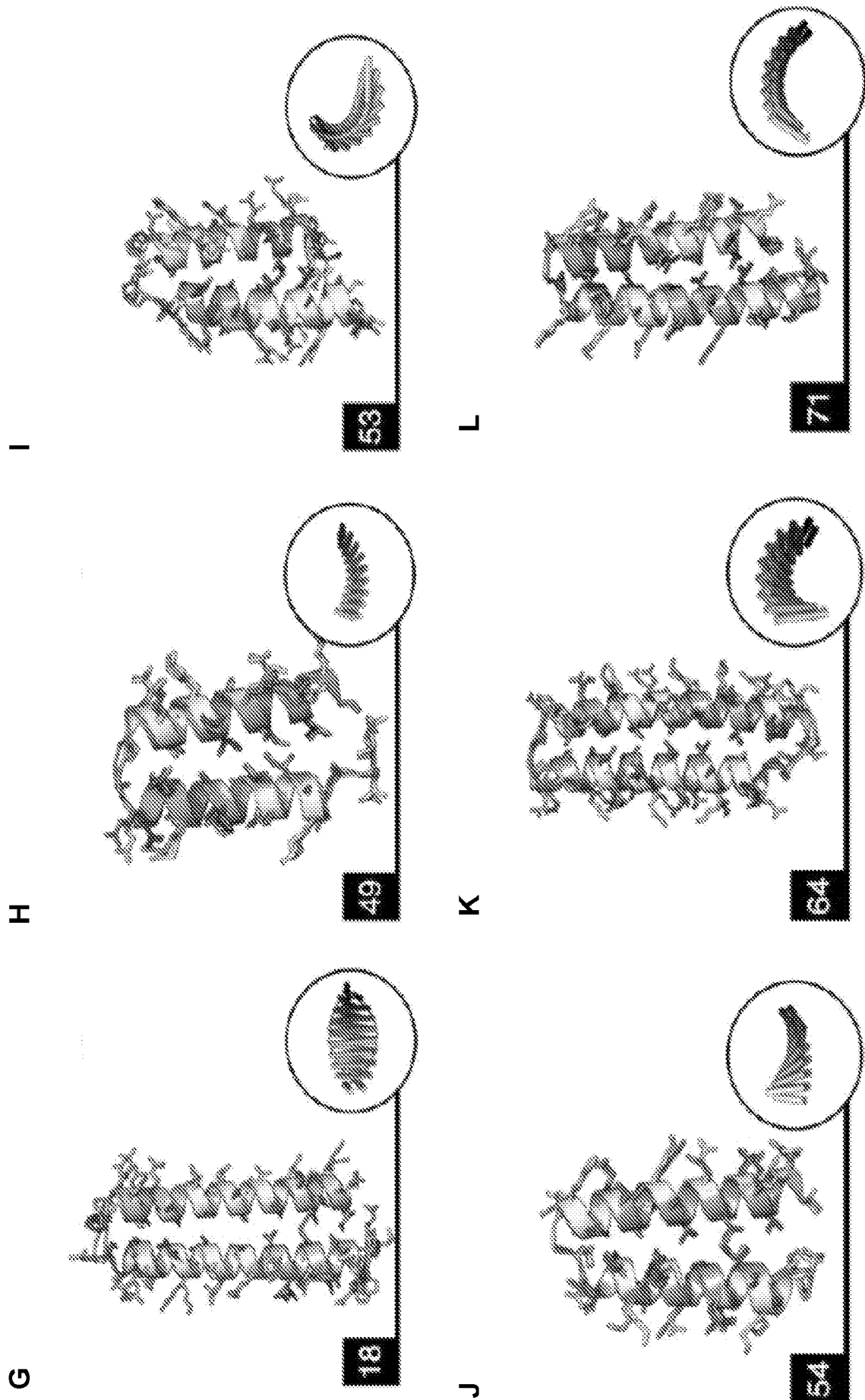


FIGURE 6G - 6L

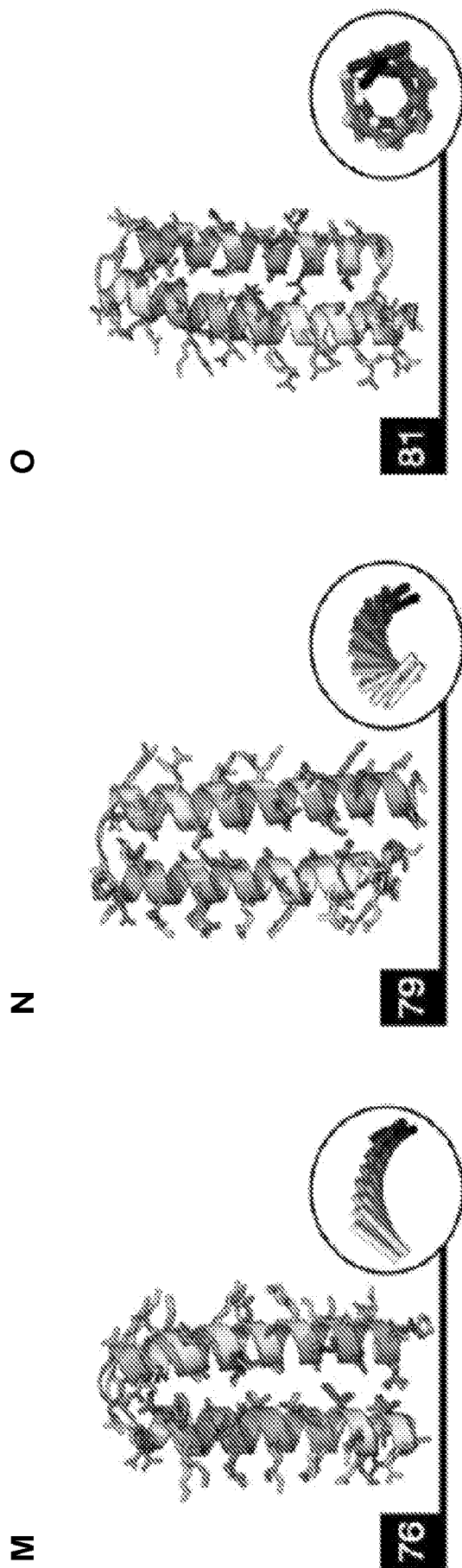


FIGURE 6M – 6O

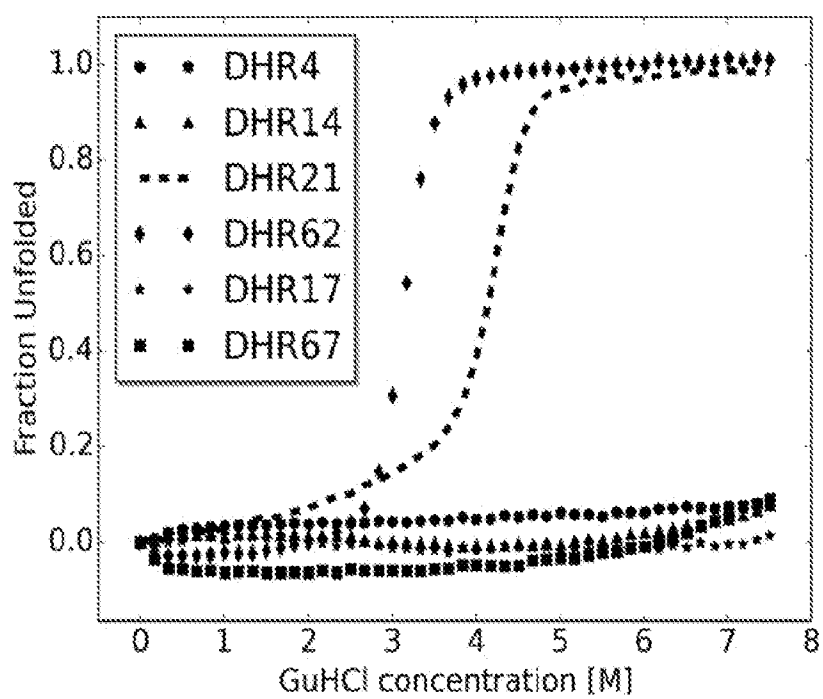


FIGURE 7

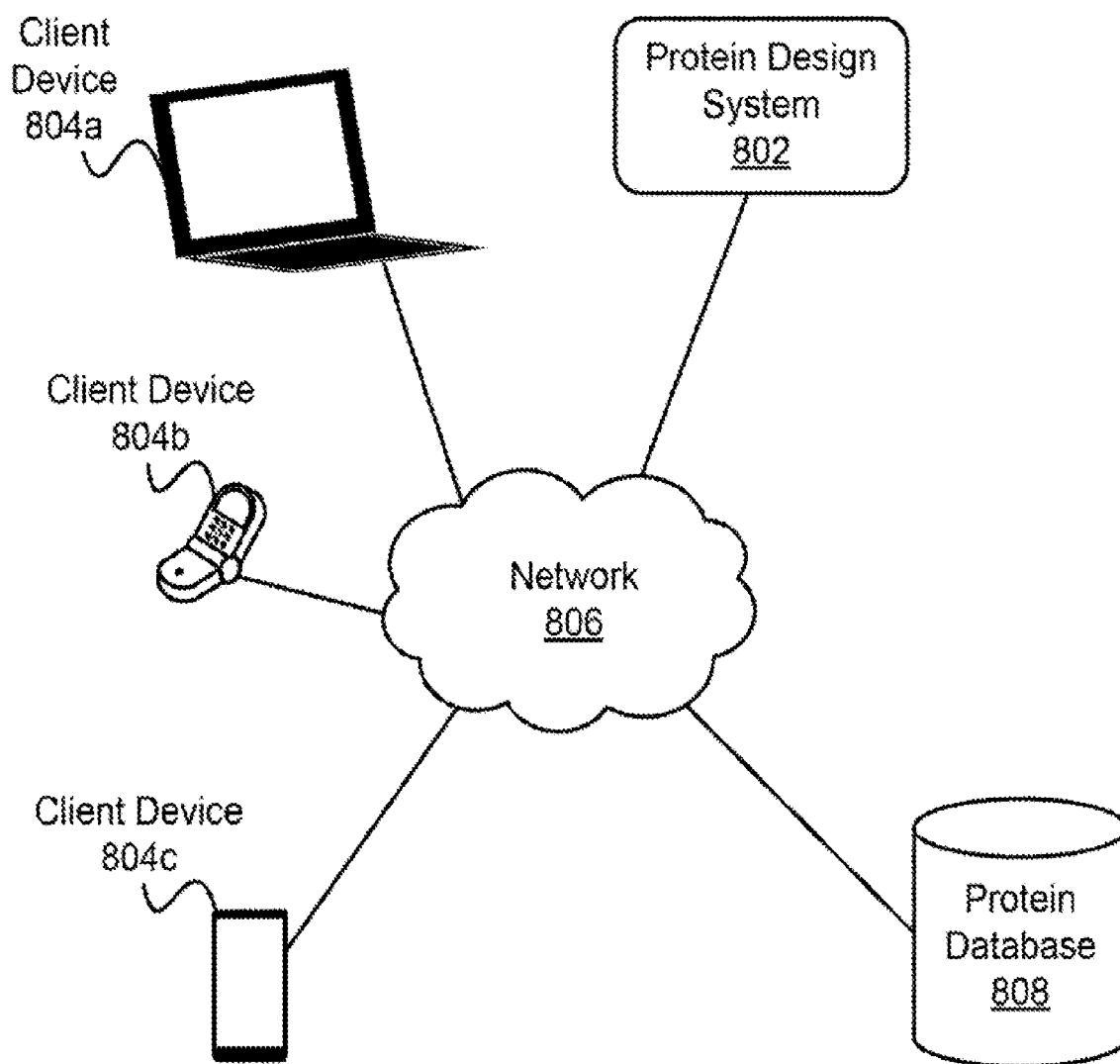


FIGURE 8

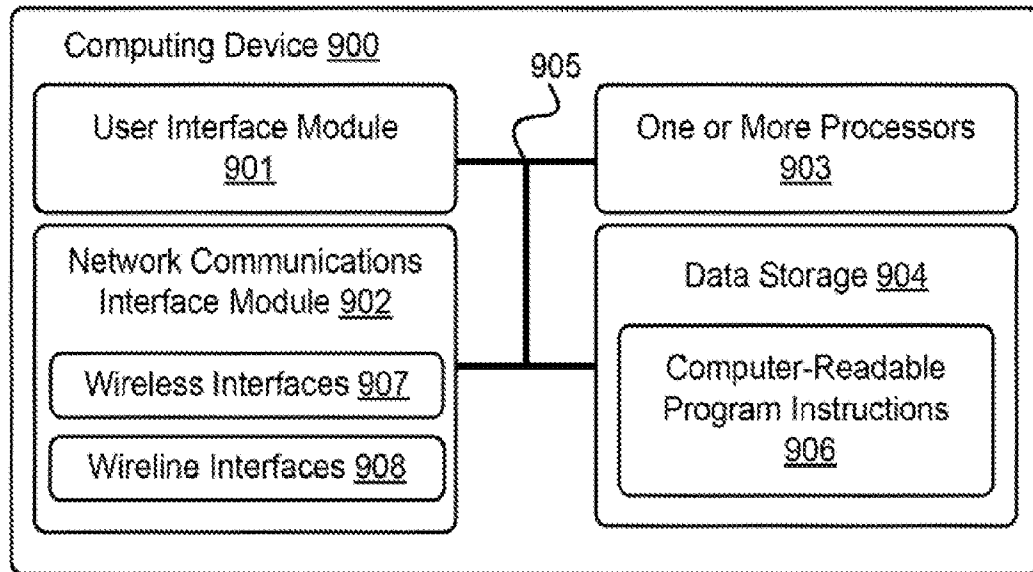


FIGURE 9A

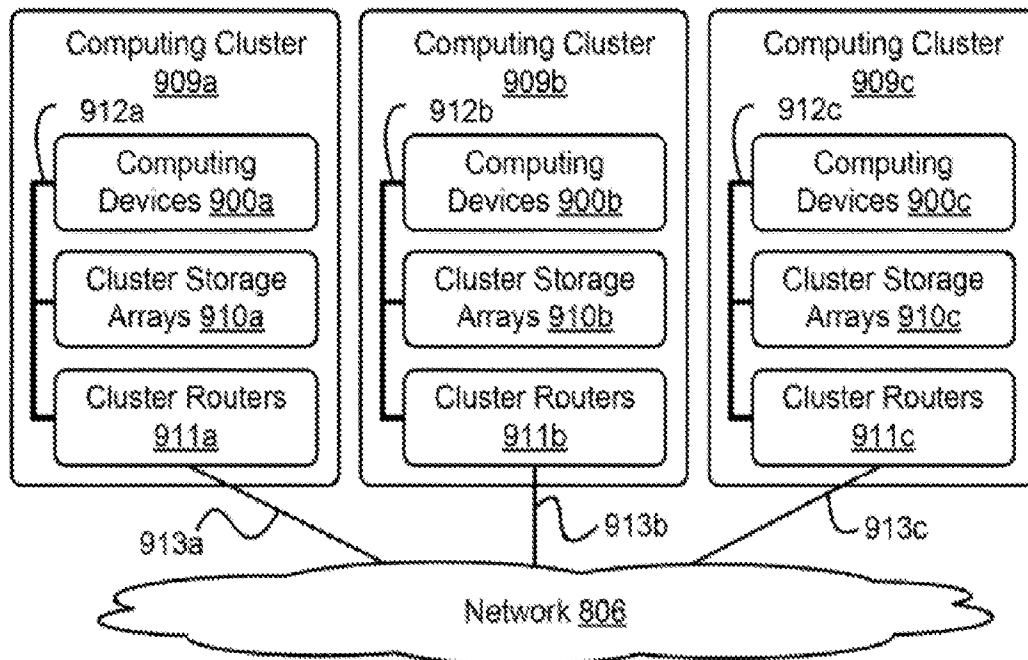


FIGURE 9B

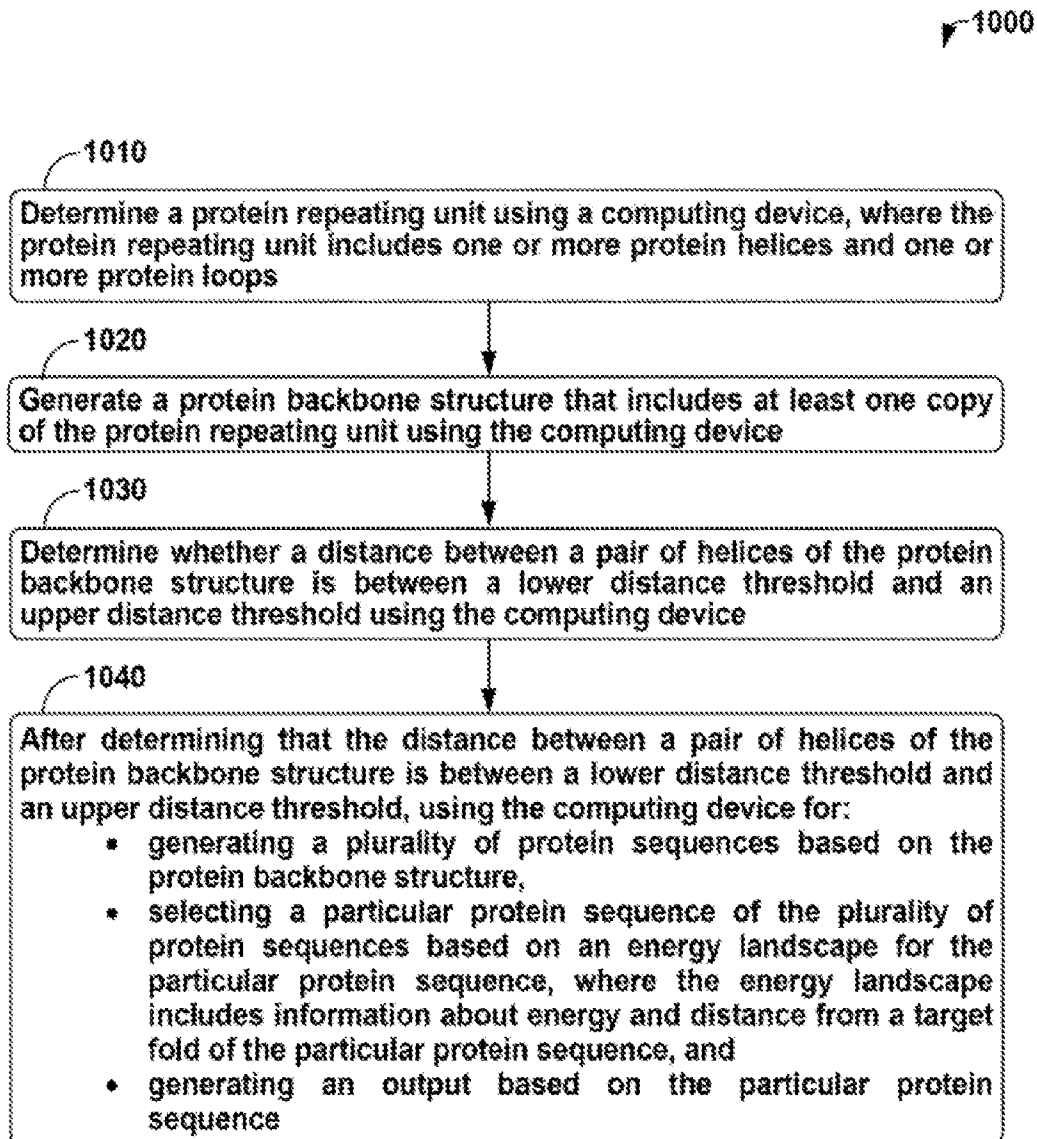


FIGURE 10