



(43) International Publication Date
3 January 2013 (03.01.2013)

(10) International Publication Number
WO 2013/003752 A2

(51) International Patent Classification:
G06F 19/16 (2011.01)

(21) International Application Number:
PCT/US2012/044996

(22) International Filing Date:
29 June 2012 (29.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/503,526 30 June 2011 (30.06.2011) US
61/548,661 18 October 2011 (18.10.2011) US

(71) Applicant (for all designated States except US): **UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION** [US/US]; 4311 11th Avenue NE, Suite 500, Seattle, WA 98105 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SCHIEF, William, R.** [US/US]; 1359 Ahlrich Avenue, Encinitas, CA 92024 (US). **CORRELA, Bruno, E.** [PT/US]; 947 1/2 Agate Street, San Diego, CA 92109 (US). **AZOITEL, Mihai** [US/US]; 401 E. Mercer Street, Apt.303, Seattle, Washington 98102 (US). **BAN, Yih-en, Andrew** [US/US]; 700 Broadway E, Apt.307, Seattle, Washington 98102 (US).

(74) Agent: **HARPER, David, S.**; 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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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, 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: METHODS FOR DESIGN OF EPITOPE SCAFFOLDS

(57) Abstract: Methods and apparatus for designing a functional protein scaffold are disclosed. A scaffold search is performed using a computing device. The scaffold search includes a search of a protein structure library for a native structure similar to a structure of a functional motif. The native structure can be supported by a protein scaffold. The native structure can be removed from the protein scaffold to create a protein sequence gap. The functional protein scaffold can be designed by inserting the functional motif into the sequence gap. The functional protein scaffold can be for targeting mAb b12.



WO 2013/003752 A2

METHODS FOR DESIGN OF EPITOPE SCAFFOLDS

RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Patent Application No. 61/503,526, entitled “Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold” filed June 30, 2011, and U.S. Provisional Patent Application No. 61/548,661, entitled “Protein Design Methods, Designed Protein Scaffolds and mAb b12 Epitope Scaffolds”, filed October 18, 2011, all of which is entirely incorporated by reference herein for all purposes.

BACKGROUND OF THE INVENTION

[0002] Unless otherwise indicated herein, the materials described in this section are not prior art to the claims in this application and are not admitted to be prior art by inclusion in this section.

[0003] Computational protein design provides valuable reagents for biomedical and biochemical research. A major limitation has been manipulation of protein scaffold structure – most computational protein design has involved sequence design on pre-determined scaffold structures or with minor scaffold backbone movement. Methods capable of sculpting protein scaffolds are needed for maximum control in the design of protein interactions and functions, but accurate scaffold backbone remodeling presents a significant challenge for computational methods owing to limited conformational sampling and imperfect energy functions.

[0004] Current systems can “graft”, or transplant structures to, the protein scaffold. Some of these structures can be defined using “functional motifs”, or sets of contiguous secondary structure elements having either functional significance to the protein or defining portions of independently folded regions. These current systems perform grafting using side-chain grafting, which relies on existing backbone conformations of proteins with known structure. Current systems have designed recognition modules, inhibitors, enzymes, and immunogens by grafting functional constellations of side-chains onto protein scaffolds of pre-defined backbone structure. However, in all these cases, the restriction to using pre-determined scaffold backbone structures limited the complexity of the functional motifs that

could be transplanted. For example, the de novo enzymes could accommodate grafting of only 3-4 catalytic groups (side-chains), whereas many natural enzymes have 6 or more, and the immunogens were limited to continuous (single segment) epitopes even though most antibody epitopes are discontinuous (involve two or more antigen segments).

[0005] Searching techniques for protein grafting have typically concentrated on searching for backbone stubs to place side chain constellations able to perform a desired function or searching for one or several contiguous segments of backbone on the protein scaffold that mimic that backbone of the functional motif to be transplanted. Typical protein design methodologies utilize fixed or nearly fixed conformations of protein scaffolds.

SUMMARY

[0006] In one aspect, a method for designing a functional protein scaffold is disclosed. A scaffold search is performed using a computing device. The scaffold search includes a search of a protein structure library for a native structure similar to a structure of a functional motif. The native structure is supported by a protein scaffold. The native structure is removed from the protein scaffold to create a protein sequence gap. A functional protein scaffold is designed by inserting the functional motif into the sequence gap of the protein scaffold.

[0007] In a second aspect, a functional protein scaffold for targeting mAb b12 is provided. The functional protein scaffold includes a protein scaffold having a first protein sequence gap and a second protein sequence gap, a primary gp120 segment positioned in the first protein sequence gap and having a primary N-terminus and a primary C-terminus, a secondary gp120 segment positioned in the second protein sequence gap and having a secondary N-terminus and a secondary C-terminus, and one or more connecting segments flanking the primary and secondary gp120 segments at each of the segment termini.

[0008] In a third aspect, the present invention provides isolated polypeptides comprising an amino acid sequence according to any of SEQ ID NOS:1-13, which can be used, for example, in the methods of the invention.

[0009] In another aspect, the present invention provides virus-like particles (VLPs) comprising the polypeptides of the invention.

[0010] In further aspects, the present invention provides isolated nucleic acids encoding the polypeptides of the invention; recombinant expression vectors comprising the isolated

nucleic acids of the invention operatively linked to a promoter; and recombinant host cells comprising the recombinant expression vectors of the invention.

[0011] In a still further aspect, the present invention provides pharmaceutical compositions, comprising the polypeptide and/or virus-like particles of the invention, and a pharmaceutically acceptable carrier.

[0012] In another aspect, the present invention provides methods for treating an HIV infection, comprising administering to a subject infected with an HIV infection an amount effective to treat the infection of the polypeptides, virus-like particles, or pharmaceutical compositions of the invention

[0013] In a further aspect, the present invention provides methods for limiting development of an HIV infection, comprising administering to a subject at risk of HIV infection an amount effective to limit development of an HIV infection of the polypeptides, virus-like particles, or pharmaceutical compositions of the invention.

[0014] In a still further aspect, the present invention provides methods for generating an immune response in a subject, comprising administering to the subject an amount effective to generate an immune response of the polypeptides, virus-like particles, or pharmaceutical compositions of the invention.

[0015] In another aspect, the present invention provides pharmaceutical composition, comprising: (a) isolated nucleic acids, recombinant expression vectors, and/or recombinant host cells of the invention; and (b) a pharmaceutically acceptable carrier.

[0016] In a further aspect, the present invention provides methods for monitoring an HIV infection in a subject and/or monitoring response of the subject to immunization by an HIV vaccine, comprising contacting the polypeptides, the VLPs, or the pharmaceutical compositions of the invention with a bodily fluid from the subject and detecting HIV-binding antibodies in the bodily fluid of the subject.

[0017] In a still further aspect, the present invention provides methods for detecting HIV binding antibodies, comprising (a) contacting the polypeptides, the VLPs, or the compositions of the invention with a composition comprising a candidate HIV binding antibody under conditions suitable for binding of HIV antibodies to the polypeptide, VLP, or composition; and (b) detecting HIV antibody complexes with the polypeptide, VLP, or composition.

[0018] In another aspect, the present invention provides methods for producing HIV antibodies, comprising (a) administering to a subject an amount effective to generate an antibody response of the polypeptides, the VLPs, and/or the compositions of the invention; and (b) isolating antibodies produced by the subject.

[0019] In a further aspect, an article of manufacture is provided. The article of manufacture includes a physical computer-readable storage medium storing instructions that, upon execution by a processor, cause the processor to perform functions. The functions include: (i) performing a scaffold search, wherein the scaffold search includes a search of a protein structure library for a native structure similar to a structure of a functional motif, and wherein the native structure is supported by a protein scaffold, (ii) removing the native structure from a representation of the protein scaffold to create a protein sequence gap, and (iii) designing a functional protein scaffold by inserting a representation of the functional motif into a representation of the protein sequence gap of the protein scaffold.

[0020] In a fourth aspect, a computing device is provided. The computing device includes a processor and data storage. The data storage stores instructions that, upon execution by the processor, cause the computing device to perform functions. The functions include: (i) performing a scaffold search, wherein the scaffold search includes a search of a protein structure library for a native structure similar to a structure of a functional motif, and wherein the native structure is supported by a protein scaffold, (ii) removing the native structure from a representation of the protein scaffold to create a protein sequence gap, and (iii) designing a functional protein scaffold by inserting a representation of the functional motif into a representation of the protein sequence gap of the protein scaffold.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] Fig. 1. Combined *in silico/in vitro* strategy for the transplantation of complex structural motifs to heterologous scaffold proteins. The discontinuous motif selected for transplantation consisted of two loops (light gray) from HIV gp120 that bind to mAb b12. Candidate scaffolds (dark gray) were identified during Scaffold Search. In Scaffold Design, the scaffold was remodeled to accommodate the transplanted motif; in Computation-Guided Library Design, directed mutagenesis libraries were created from an ensemble of low energy designs with diverse structures and sequences in the regions connecting the motif to the scaffold; in *in vitro* Screening, libraries were screened sequentially by yeast display to isolate scaffold variants with high affinity for b12.

[0022] Fig. 2. Isolation of scaffold 2bodx variants with high b12 affinity and specificity. (A) Screening of the computation-guided libraries led to rapid enrichment of clones with high b12 affinity; R1-R3 refer to rounds 1-3 of selection. (B) SPR equilibrium analysis of the initial computational design (2bodx_03) and the 2bodx variants identified from the directed libraries (see Figures 12A and 12B). (C) 2bodx_43 binds to b12, but not to CD4 or other antibodies that target the CD4 binding site on gp120;

[0023] Figures 3A-C. Atomic-level recapitulation of the b12-gp120 interface on the b12-2bodx_43 complex. Figure 3A shows an overall structure of b12 in complex with 2bodx_43.

[0024] Figure 3B shows the conformation of the transplanted loops (light gray) in 2bodx_43 (darkest gray) accurately mimics their conformation on gp120 (medium gray) (0.8 Å and 1.5 Å backbone rmsd for CD4b and ODe loop respectively).

[0025] Figure 3C shows the epitope and paratope take highly similar conformations in the two complexes; conformations of side-chains (sticks) making important contacts in the b12-gp120 complex are preserved at the 2bodx_43-b12 interface; H1, H2 and H3 refer to the CDR loops of b12.

[0026] Figures 4A-C show the Multigraft Match algorithm, with Figure 4A showing an example Multigraft Match workflow, Figure 4B showing example alignment schemes available in Multigraft Match, and Figure 4C showing examples of the atoms used to perform the structural alignments and the chain-break measurements.

[0027] Figure 5 is a schematic representation of the Multigraft Design workflow. The cartoons on the right provide an example from the design of the b12 epitope-scaffold, with the epitope in light gray and the native scaffold in darker shades of gray.

[0028] Figure 6A shows a sequence alignment of the native 2bod protein and engineered 2bodx variants; the “2bodx_03” residues in bold were introduced in the initial computational design; residues in 2bodx_43 and 2bodx_45 were identified from the computation-guided libraries (Libraries 1, 2, and 3) and the random mutagenesis library. Stars (*) indicate regions of missing density in the native 2bod structure. Symbols above the residue positions indicate the sampling freedom during the sequence-structure diversification stage: ○ – positions with backbone flexibility for which the library allowed multiple amino acids from computational design; ◇ – positions allowed backbone flexibility for which the library was confined to the native 2bod and the 2bodx_03 identities; ● – positions with fixed

backbone conformation for which the library was confined to the native 2bod and the 2bodx_03 residue identities; ♦ - positions with fixed backbone conformation for which the library allowed multiple amino acids from computational design.

[0029] Figure 6B shows locations of the sequence changes mapped on the crystal structure of 2bodx_43.

[0030] Figure 6C shows structural alignment of the crystal structures of native 2bod (medium gray) and 2bodx_43 (dark gray); epitope loops are shown in light gray;

[0031] Figures 7A-C show interaction of 2bodx_03 with b12 mAb, with Figure 7A showing 2bodx_03 bound b12 at high concentrations by SPR, Figure 7B introducing the D114R mutation that eliminated the binding signal, and Figure 7C showing SPR analysis of 2bodx_03-b12 interaction indicates a $K_D \approx 300 \mu\text{M}$;

[0032] Figure 8 shows an SPR analysis of the interaction between mAb b12 and the 2bodx variants isolated from the experimental libraries.

[0033] Figures 9A-E. Computational protocol 900 for structure-sequence diversification including conformational ensembles of epitope connecting loops 910 shown in Figure 9B, designing and filtering 920 shown in Figure 9C, explicit recombination of designs 930 shown in Figure 9D, designing and filtering (second pass), conformational resampling of interacting segment(s) 940 shown in Figure 9E, and designing and filtering (third pass).

[0034] Figures 10A and 10B. Amino acid frequency at 21 positions based on the sequence alignment of the top 45 models generated by the computational structure-sequence diversification procedure. Positions shown in with diagonal-lined bars were designed during the computational procedure; positions shown in cross-hatched bars (*) were fixed at their respective 2bodx_03 identities during the computational procedure but were allowed to mutate en bloc in the experimental libraries between the computational design and the wild-type 2bodx. Only the residue variants shown in diagonal-lined and cross-hatched color bars were included in the experimental libraries. Figure 10A shows sequence alignment of residues adjacent to the ODe loop and Figure 10B shows sequence alignment of residues adjacent to the CD4b loop.

[0035] Figure 11A shows that libraries were constructed and transformed by taking advantage of homologous recombination in yeast.

[0036] Figure 11B shows theoretical and experimental size of the investigated libraries.

[0037] Figures 12A and 12B show sequence diversity of the three experimental libraries. Library 1, shown in Figure 12A, explored all the computation-generated sequence diversity for the ODe loop combined with the CD4b loop variants from 8 of the top 45 computational models; Library 2, shown in Figure 12B, explored all the computation-generated sequence diversity for the CD4b loop together with the ODe loop variants isolated after three selection rounds of Library 1; 48 clones from the third selection round of Library 1 were analyzed and 18 unique ODe loop sequences were identified; C. Library 3, shown in Figure 12A, explored sequence variants not tested in Libraries 1 and 2. In both Figures 12A and 12B, residues in bold correspond to the residues from the selected 2bodx_42; residues shown with accent marks, e.g., M', were introduced in the experimental libraries due to codon degeneracy; residues shown with stars, e.g., T*, were changed in the initial 2bodx_03 design, but were not in direct contact with the transplanted loops and as such allowed to keep 2bodx_03 identity or revert to their native 2bod identity; the sequence of 2bodx_03 is shown as reference.

[0038] Figure 13. Summary of library screening results. Frequency refers to the abundance of one clone within the total clonal population at a given selection round; ODe_loop_42 refers to the clone from Library 2 that had the same ODe loop sequence as 2bodx_42.

[0039] Figure 14. mAb b12 interacts with 2bodx_43 specifically. D114R mutation eliminates binding of b12 to 2bodx_43; 2bodx_43 and 2bodx_43 D114R at 1 μ M were injected over mAb b12 captured on α -hIgG amine-coupled to a CM5 chip.

[0040] Figure 15. Biophysical characterization of 2bodx_43. Far-UV Circular Dichroism shows a characteristic spectrum of a folded α/β protein. Thermal denaturation shows cooperative unfolding with a melting temperature of $T_m \approx 75$ °C. Size exclusion chromatography (SEC) and static light scattering (SLS) show a monomer in solution with the expected molecular weight of ≈ 30 kDa.

[0041] Figures 16A-D. Contribution of the individual residues selected during *in vitro* evolution to the b12 affinity of 2bodx_44. 16 single-mutant constructs where “evolved” 2bodx_44 residues were individually reverted to their corresponding 2bodx_03 identity and were analyzed for b12 binding by FACS on the surface of yeast. Figure 16A shows binding levels of single-mutant constructs at 1 μ M, 100nM and 10nM b12 Fab; reported values are given as a fraction of the 2bodx_44 binding level.

[0042] Figure 16B shows residues in which mutations had large effect on b12 binding with labels. Residues with reduced or no effect upon mutation are not labeled.

[0043] Figure 16C shows display levels of single-mutant constructs on the surface of yeast expressed as a fraction of the display level of 2bodx_44; the effect of two 2bodx_44 to 2bodx_03 reversions was not analyzed (L122W and V118A); A118V increased the b12 affinity of 2bodx_03 by approximately 10-fold and that of 2bodx_42 by 5-fold.

[0044] Figure 16D shows b12 affinity of a 2bodx_03 construct with 9 of the mutations present in 2bodx_44 (N72D, D74G, G109L, G110Y, A121W, W122L, F124T, A118V, A125V); 6 of the mutations were shown in Figure 16A to have the largest effect on the binding of 2bodx_44 (N72D, D74G, G109L, G110Y, A121W, F124T; two of the mutations were not analyzed in A) (L122W and V118A), but were included in case they affected b12 binding; the recombinantly expressed construct had a KD of 1.5 μ M based on kinetic and equilibrium analysis of SPR data.

[0045] Figure 17. Comparison of the b12 footprint on the structures of gp120 (left and partially represented) and 2bodx_43 (right). The antibody footprints in the epitope region are in a light gray in the top-central portion of the region, and the extra-contacts observed in the 2bodx_43-b12 complex are colored in light gray in the central portion of the complex; b12 heavy chain is shown in black in the top-left portions of the top figures for both gp120 and 2bodx, and the light chain is shown in the top-right portions of the top figures for both gp120 and 2bodx.

[0046] Figure 18. Interaction of 2bodx_43 with mAb b13. b13 and b12 mAbs were each captured on anti-human IgG antibodies surfaces amine-coupled to a CM5 chip. 2bodx_43 was then injected at 96 μ M and 320 μ M. Significant responses against mAb b13 (~15% of the theoretical Rmax) were observed at 320 μ M 2bodx_43; binding responses to mAb b12 were close to the theoretical Rmax at both 2bodx_43 concentrations. gp120 was injected as a positive control; higher overall response levels for gp120 binding were due to its higher molecular weight.

[0047] Figure 19 is a block diagram of an example computing network.

[0048] Figure 20A is a block diagram of an example computing device.

[0049] Figure 20B depicts an example cloud-based server system.

DETAILED DESCRIPTION

[0050] All references cited are herein incorporated by reference in their entirety. Within this application, unless otherwise stated, the techniques utilized may be found in any of several well-known references such as: *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, CA), "Guide to Protein Purification" in *Methods in Enzymology* (M.P. Deutscher, ed., (1990) Academic Press, Inc.); *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, 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).

[0051] As used herein, the amino acid residues are abbreviated as follows: alanine (Ala; A), asparagine (Asn; N), aspartic acid (Asp; D), arginine (Arg; R), cysteine (Cys; C), glutamic acid (Glu; E), glutamine (Gln; Q), glycine (Gly; G), histidine (His; H), isoleucine (Ile; I), leucine (Leu; L), lysine (Lys; K), methionine (Met; M), phenylalanine (Phe; F), proline (Pro; P), serine (Ser; S), threonine (Thr; T), tryptophan (Trp; W), tyrosine (Tyr; Y), and valine (Val; V).

[0052] 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.

[0053] All embodiments of any aspect of the invention can be used in combination, unless the context clearly dictates otherwise.

[0054] The present disclosure is generally directed to functional protein scaffold design methods, designed protein scaffolds and mAb b12 epitope scaffolds. In particular, protein grafting can be used to both to test protein structure manipulations and for translational applications. Aspects of the present disclosure are directed to a method which integrates computational design with experimental selection for grafting single- or multi-segment functional motifs onto protein scaffolds, regardless of the starting protein scaffold. In some embodiments, the method includes changing, inserting or removing an amino acid connecting segment flanking the function motif. In another embodiment amino acid changes can be made within the single- or multi-segment functional motifs.

[0055] Aspects of the present disclosure include a computational-based scaffold design method for designing protein scaffolds to stabilize functional sites. The scaffold design method can be used to design unique scaffolds of any type which can be applied generally to stabilize functional or binding sites on proteins including but not limited to antibody epitopes or enzyme active sites. One exemplary embodiment of the present disclosure is directed to designing and synthesizing a functional protein scaffold containing a functional motif and connecting segments for the HIV epitope gp120, targeted by the cross-neutralizing antibody b12. Those of skill in the art will recognize that the methods can be used to design functional protein scaffolds of any desired activity.

[0056] The methods integrate computational design with experimental selection for grafting the backbone and side-chains of single- or multi-segment structural motifs onto scaffold proteins. Advantages for backbone grafting are that backbone grafting enables grafting of functional sites for which there are no suitable matches for side-chain grafting, and backbone grafting can enable improved functional activity compared to side-chain grafting even when there are matches for side-chain grafting. The ability to generate highly specific binding agents by backbone grafting can be useful in a variety of contexts including, for example, design of enzymes, inhibitors, binding reagents, therapeutics, and vaccines.

[0057] The disclosed flexible backbone computational design enables an efficient search through a large conformation and sequence space in which exhaustive exploration was experimentally unfeasible. The results demonstrate that computation-guided library design is a powerful complement to other molecular evolution methodologies for rational manipulation of protein backbone structure to control interaction and function. One of ordinary skill in the art will recognize that the approach described here can be used to design additional hybrid protein structures having desired interactions and functions.

[0058] Certain embodiments described herein enable generation of specific binding agents utilizing computational and experimental techniques to graft functional motifs, as either single or multiple segments onto unrelated protein scaffolds. In one embodiment, the combined *in silico/in vitro* approach was successful for transplantation of the backbone and side-chains of a discontinuous binding motif to a protein scaffold. This method was used to transplant the two-segment HIV gp120 epitope onto an unrelated scaffold. The designed functional protein scaffold was endowed with high affinity and specificity for mAb b12, and its binding to b12 recapitulated the gp120-b12 interaction with high structural fidelity. These

features indicate that such functional protein scaffolds have promise as tools for HIV vaccine research as well as for other bNAbs such as VRC01 that binds an overlapping but different epitope on gp120.

[0059] Important reasons to graft functional motifs onto scaffolds are: (a) Structural stabilization of the active conformation of a binding or functional site to improve binding or functional activity; here the functional motif is embedded into rigid scaffolds to stabilize the desired conformation(s) of the motif to improve activity; we note that the motif may be naturally occurring, in which case transplantation to and rationally embedding within a rigid artificial scaffold allows functional optimization in the absence of any evolutionary pressures or other biological constraints that impinge on natural proteins; the motif also may be artificially designed or predicted (such as a theozyme model for an enzymatic transition state computed by quantum mechanical or other methods), in which case a natural protein host may not be available so embedding the motif in an artificial scaffold may be essential to stabilize the functional conformation (b) transferring a binding or functional site to one or more scaffolds in which the scaffold(s) itself provides additional advantages such as (i) the scaffold contains one or more additional functional sites (ii) improved protein expression or ability to produce the scaffold in alternative expression systems such as E.coli, (iii) improved protein solubility (iv) improved protein stability against thermal or chemical denaturation (v) smaller scaffolds to reduce cost of production and/or reduce immunogenicity if immune responses are not desired (vi) for applications where immune responses to the epitope are desired: improved immunogenicity of the functional site – for example by using a smaller scaffold to reduce potentially distracting epitopes, or by using a larger scaffold with additional CD4 T help epitopes, or using a scaffold that can be glycosylated or PEGylated, or by using a multivalent scaffold or using a series of scaffolds that can be combined in cocktails and/or sequences in immunization regimens to focus immune responses on the functional epitope, or by using scaffolds that can be more easily incorporated into existing vaccines such as live attenuated vaccines or virus-like particle vaccines.

[0060] These techniques allow for more aggressive manipulation of sequence and structure (and function) of pre-existing protein scaffolds, opening unprecedented avenues for the functional repurposing of protein scaffolds for novel biological functions. These techniques can be used to generate a large number of sequences that can be screened with a high-throughput methodology such as yeast display. The “Multigraft Match” matching algorithm uses a definition of structural compatibility where the algorithm searches for pairs

of residues which have similar distances to the end-point distances of the motif to be grafted. The Multigraft Match algorithm allows for a large number of combinations of matching schemes for different segments that compose a potential functional motif.

[0061] The protein sampling component disclosed herein builds the segments that connect the scaffold to the functional motif using loop modeling strategies and followed by sequence design to find low-energy sequences for the novel conformations. Incorporating backbone flexibility modeling into the grafting design process allows improved activity and expanded functionality from complex designs. Towards this end, we developed a hybrid computational-experimental method for grafting the backbone and side-chains of functional motifs or epitopes onto scaffolds.

[0062] Figure 1 shows a method for designing a functional protein scaffold. The method can include (i) performing scaffold search 110 using a computing device, where the search can include a search of a protein structure library for a native structure similar to a structure of a functional motif, and where the native structure is supported by a protein scaffold, (ii) removing the native structure from the protein scaffold 121 to create a protein sequence gap; (iii) designing a functional protein scaffold by inserting 122 the functional motif into the sequence gap of the protein scaffold. The method can also include changing, inserting or removing an amino acid connecting segment flanking the functional motif or flanking individual segments of a multi-segment functional motif. One of ordinary skill in the art will recognize that changing an amino acid segment can include modifying or replacing one or more amino acids within the segment, inserting additional amino acids within the segment or removing one or more amino acids within the segment. One of ordinary skill in the art will also recognize that flanking can include embodiments wherein a connecting or other amino acid segment is on one side or another of the motif or flanking can refer to embodiments in which a connecting or other amino acid segment is found on both sides the motif. In one embodiment, the amino acid structure of the functional motif can also be changed (e.g., modifying, replacing, inserting, removing one or more amino acids).

[0063] The method can further include creating a computation-guided library 130 accessible by the computing device, where the computation-guided library includes a plurality of library entries, and wherein at least one library entry of the plurality of library entries relates to the functional protein scaffold. In some embodiments, a unique library entry may contain only a single amino acid change from an original structure of the functional

protein scaffold. In other embodiments, a unique library entry may contain more than one amino acid change from an original structure. In some embodiments a set of mutagenesis libraries are selected from an ensemble of designs with expanded sequence diversity in any one of the segments of the functional motif, the connecting segments or the protein scaffold. The method can further include performing *in vitro* screening 140 of the unique library entries to identify clones with desired functional activity.

[0064] In one embodiment, this method was used to graft a discontinuous HIV gp120 epitope, targeted by the broadly-neutralizing antibody b12 (21), onto an unrelated scaffold. b12 binds to a conserved epitope within the CD4- binding site of gp120 (22), an area of great interest for vaccine design. In this embodiment, we focused on transplantation of two segments from gp120 – residues 365-372 and 472-476, known as the ‘CD4 binding (CD4b) loop’ and ‘Outer domain exit (ODE) loop’, respectively. The final scaffolds bound b12 with specificity and with similar affinity as gp120, and crystallographic analysis of a scaffold bound to b12 revealed structural mimicry of the gp120-b12 complex structure. The results from this embodiment demonstrate that combining computational and selection methods allow precise protein structure molding, and show how functional motif grafting may enable design of other functional proteins.

[0065] In one embodiment, a functional protein scaffold for targeting mAb b12 includes a protein scaffold having a first protein sequence gap and a second protein sequence gap. The functional protein scaffold also includes a primary gp120 segment positioned in the first protein sequence gap and having an N-terminus and a C-terminus, and a secondary gp120 segment positioned in the second protein sequence gap and having an N-terminus and a C-terminus. The functional protein scaffold further includes connecting segments flanking the primary and secondary gp120 segments at each of the motif termini.

Scaffold Search:

[0066] The scaffold search can include a search of a protein structure library, such as a Protein Data Bank (PDB) shown in Figure 4A. for a native structure similar to a structure of a functional motif, and wherein the native structure is supported by a protein scaffold. Each functional motif can include one or more segments. For each segment, two “insert positions” or locations on the protein scaffold to attach the segment are needed – one insert position to add the beginning of the segment to the protein scaffold and one insert position to add the end of the segment to the protein scaffold. So, for a functional motif with S segments, $2S$

insertion points need to be found on the protein scaffold. In some embodiments, the scaffold search can further search for superposition matches for grafting or more segments to side chains of the backbone.

[0067] For all possible combinations of 2*S* insert positions in every candidate scaffold, Multigraft Match can predict at low-resolution whether the *S* segments of an input functional motif can be grafted onto the protein scaffold while maintaining backbone continuity and avoiding steric clash. In one scenario, eleven candidate scaffolds satisfied the geometrical and steric clash requirements in the matching stage for a two-segment functional motif and were selected for design.

Primary Loop Matching

[0068] A “Multigraft Match” algorithm, shown in Figures 4A-C, can search a culled protein structure database. In some embodiments, the protein structure database can store a large number, *e.g.*, approximately 30,000, of protein chains for suitable scaffolds for a functional motif. To cull the protein structure database, protein chains can be required to be (a) X-ray crystal structures having a resolution better than a certain threshold; *e.g.* 3.0 Å and (b) without small molecule ligands. Other culling criteria are possible as well. In some embodiments, the PDB discussed above can act as the protein structure database.

[0069] To begin Multigraft Match, each segment of the functional motif is marked as unfound. In Multigraft Match, one of the remaining unfound segments of the functional motif is selected as the Primary Loop, as shown in Figure 4A. Once a match for the Primary Loop is found, the segment selected as Primary Loop is marked as found a remaining unfound segment is then selected as a new Primary Loop, until all segments of the functional motif are marked as found.

[0070] The Primary Loop can be aligned on the residues of the scaffolds contained in a culled protein structure database to identify potential sites where it can be transferred as shown in Figure 4A. Several types of alignments can be utilized to cover a variety of motif grafting scenarios: full backbone superposition (Superposition), C-terminus alignment (C2N), N-terminus alignment (N2C) and End-point alignment (E). Examples of these alignments are shown in Figures 4B and 4C.

[0071] The superposition alignment portion of the Primary Loop matches the Primary Loop against segments in the query scaffolds of the protein structure database to identify regions with close structural mimicry. From the four alignment systems available only

superposition evaluates the local similarity of the existing backbone comprehensively. Superposition is shown pictorially in Figures 4B. The C-terminus and N-terminus alignments are conceptually identical, differing only by which terminus of the Primary Loop is aligned onto the scaffold (Fig. 4B).

[0072] Within the C2N and N2C alignments several subsets of three atoms can be used to perform the structural alignments: N centered, shown in Fig. 4C with a C2N_N label, C centered, shown in Fig. 4C with a N2C_C label, and C α centered, shown in Fig. 4C with an End-Point match label. The End-point alignment (Fig. 4B) superimposes the three backbone atoms (N, C α , C) of both Primary Loop termini onto the corresponding atoms of residue pairs on the query scaffold as shown in Figures 4B and 4C.

[0073] To improve speed, Multigraft Match can use filters to identify a physically reasonable match between a motif and the scaffold. For example, suppose the Primary Loop, which is matched first as mentioned above, is not compatible with a given location in the scaffold according to a filter. In this example, the match is then stopped and a next scaffold location is attempted. This use of filters saves computations and time that could be used in matching all the remaining segments of the motif to be transplanted / grafted.

[0074] Figure 4A shows two different kinds of filters are assessed after Primary Loop alignment: geometric and steric clash. The geometric filter for the superposition alignment evaluates the backbone root mean square deviation (rmsd) of the Primary Loop relative to a segment in the scaffold. In the other three alignment types (C2N, N2C, E) the geometric filter also evaluates the rmsd between the termini of the Primary Loop and proximal scaffold residues to yield a chain-break score; the subset of atoms used to compute the chain-break is also shown in Figure 4C. The calculated rmsd value can be determined to indicate a match if the calculated rmsd value is less than a predefined threshold rmsd value.

[0075] The steric-clash filter ensures that the particular alignment of the Primary Loop on the scaffold does not clash with other backbone regions of the scaffold (inter-clash). Steric clash can be measured by the van der Waals repulsive interaction assessed in the all-atom energy function. In the Rosetta protein modeling software, steric clash can be called “fa_rep”, for an amount of energy due to the Lennard-Jones repulsive force. In some embodiments, steric clash can be measured by any other techniques that measure of atomic overlap.

[0076] The calculated steric-clash value for the Primary Loop can be determined to indicate if the calculated steric-clash value is less than a predefined threshold. Figure 4A

shows that Primary Loops only match when both the Geometric and Steric Clash filters are satisfied within user-defined thresholds

[0077] In other embodiments, more, fewer, and/or different filters can be used. For example, a secondary structure agreement filter can filter matches by assessing agreement between the functional motif segments and the scaffold segments, a directional filter can filter matches by assessing the angle computed between the stubs of the scaffold and the motif to be grafted, and/or a residue deletion filter can filter matches by assessing how many residues are being deleted from the scaffold. Other filters can be used as well or instead.

[0078] In some embodiments, the Primary Loop alignment determines the rigid body orientation of the Secondary segments and of the associated binding partner and this orientation is reconstituted at later stages of the procedure. Other embodiments can begin with this same rigid body alignment and then allow limited motion of Secondary Loop(s) relative to the Primary Loop. Still other embodiments can enable changing the rigid-body orientation of the associated binding partner, perhaps by a rigid-body minimization or by limited docking.

Secondary Segment Matching

[0079] After satisfying the Primary Loop matching filters, Figure 4A shows that the remaining segments, or Secondary Loops/segments, of the input complex motif are recovered in the context of the putative scaffolds while retaining an original orientation to the Primary Loop. Figure 4A shows that, for each Secondary Loop to be added onto the primary match, a search for scaffold residues located within the proximity of the Secondary Loop can be performed. The criteria to select a Secondary Loop are those of the End-point match described for the Primary Loop. Searching for scaffold residues using an end-point match can reduce the number of scaffold residues to query, thereby increasing the speed and efficiency of the computational procedure. Figure 4A then shows that Secondary Loop alignment has to satisfy both the geometric and steric clash filters. If both criteria are satisfied, Multigraph Match proceeds to the next stage.

Clash check

[0080] Figure 4A shows that, after Primary Loop and Secondary Loop matching, Multigraph Match recapitulates the orientation of an input complex between a binding partner and a “hybrid” scaffold structure composed of the scaffold with the input functional motif in the match-derived orientation. The theoretical binding energy of the input complex is

computed and, as shown in Figure 4A, matches with steric clashes across the interface are discarded. For example, a computed theoretical binding energy can be compared to a predefined threshold steric-clash value, and if the theoretical binding energy exceeds the threshold steric-clash value, the input complex is rejected. In some implementations, in the Rosetta programming language, the threshold steric-clash value can be 10 or 100 Rosetta energy units, in which the energy refers only to the binding energy ($E_{\text{complex}} - E_{\text{partner1}} - E_{\text{partner2}}$) computed by the `fa_rep` term.

Scaffold Design Using the Multigraft Design Algorithm:

[0081] The scaffold design can includes steps wherein the functional motif replaces native scaffold segments and connecting segments, and surrounding side-chains are designed to support the functional motif conformation. In one embodiment, the “Multigraft Design” algorithm can be used to design a functional protein scaffold. Multigraft Design algorithm takes a preliminary rigid-body orientation for a discontinuous epitope relative to a native scaffold as an input, and can delete appropriate regions of the native scaffold, build new segments to connect the epitope to the native scaffold, and generate side-chains neighboring the epitope and connecting segments to support the graft (Fig. 5). This involved structural manipulations including replacement of ordered secondary structure motifs by the functional motif segments, flexible backbone modeling of two or more connecting segments, and sequence design of 10 or more core residues. Several design variants of each candidate scaffold can be tested for expression and purification in *E. coli*.

[0082] In some embodiments, the Multigraft Design algorithm can be implemented using the Rosetta language, while in other embodiments, the Multigraft Design algorithm can be using one or more other protein design packages, such as MSL or Orbit. In particular, the Rosetta platform includes many crystal structures of proteins designed using Rosetta that validate the energy function and conformational sampling methods.

[0083] Figure 5 shows the Multigraft Design algorithm, which is composed of several stages. These stages can include, but are not limited to: I) match reconstitution; II) building connecting segments by fragment insertion; III) sequence design; IV) filtering; and V) human-guided design. In one embodiment, Multigraft Design was implemented to automatically receive, as an input, Multigraft Match output to create a streamlined application using both Multigraft Match and Multigraft Design stages, and thus keep track of the mapping between the matched motif residues and the scaffold. In some embodiments, stages

II and III can be carried out using RosettaRemodel functionality that uses the same Rosetta algorithms as Multigraft Design to sample conformation and sequence space. Below are brief descriptions of each stage of Multigraft Design.

Match reconstitution

[0084] A hybrid structure of the protein scaffold with the functional motif in the matched rigid-body orientation can be assembled during match reconstruction. Figure 5 shows that segments of the protein scaffold can be swapped with the functional motif as previously defined in the matching stage.

Generation of connecting segments

[0085] Figure 5 shows that the hybrid structures can have one or multiple backbone discontinuities where the backbone of the protein scaffold and the functional motif are disconnected. To reconnect the polypeptide chain and generate continuous backbone conformations one can use a backbone building algorithm that performs fragment insertion and Cyclic Coordinates Descent. The backbone length and secondary structure of the segments to be built can be user-defined. In some embodiments, different connecting segment lengths and secondary-structure combinations can be attempted in order to maximize the probability of obtaining a properly connected backbone. The backbone conformations able to satisfy the chain-break filter can be carried to the following stages.

Sequence design

[0086] In the sequence design stage, as shown in Figure 5, the residues from the newly rebuilt segments and neighboring regions are allowed to change their identity in order to find low energy sequences for the newly generated backbone conformations. Following the initial sequence variation, the connecting segments can be subjected to multiple rounds of side-chain and backbone minimization. The steps in this stage can be carried out using a full-atom energy function for accurate energetic assessment.

Filtering

[0087] Figure 5 shows that the filtering stage uses one or more filters to rank and select designs with lower Rosetta energy that exhibit native-like protein features. The filters can identify designs with structural flaws such as outlier backbone dihedral angles according to the Ramachandran distributions; hollow cavities, and large numbers of buried unsatisfied polar atoms.

Human-guided design

[0088] The focus of this stage, shown in Figure 5 as a “Final Design Stage”, is to inspect and correct aspects that were not considered in the automated design stage. Interventions can include removing unpaired cysteines, functional sites, oligomerization interfaces and mutating solvent exposed residues to enhance the protein solubility.

[0089] This stage is optional, but can be used to improve the frequency of successful designs at the experimental testing stage. Some final design stage interventions, can be and/or are computerized; *e.g.*, removing unpaired cysteines and redesigning hydrophobic exposed surfaces for improved solubility. Other interventions can be partially or fully automated, such as redesigning oligomerization interfaces.

Computation-guided library design

[0090] Computation-guided library design includes creating a set of mutagenesis libraries selected from an ensemble of designs with expanded sequence diversity in the connecting segments.

[0091] Figure 9A shows a structure-sequence diversification protocol 900 that can be employed for computation-guided library design. To ensure fine conformational sampling, ensembles of backbone conformations are generated for each connecting segment separately, as shown in Figure 1. The ensembles are subjected to sequence design and filtering to identify mutations that stabilize the different backbone conformations; several of the lowest energy models for each segment are recombined *in silico* to generate all possible combinations; the recombined models are also subjected to sequence design, resampled, and then filtered based on several Rosetta scores including total energy, rama (penalizes unfavorable backbone dihedral angles), packing, and buried unsatisfied polars.

[0092] Figures 10A and 10B show that the best models according to the filters are used to generate lists of allowed residues for the positions in the connecting segments. The diversity at each position can be further reduced by eliminating residues that occurred at low frequency, that were similar in size and chemical nature to more frequent residues, or that were judged likely to bury a polar side chain. Additional library diversity can be introduced by the degenerate codons used for library construction. Computation-guided library design reduces the number of sequences to screen experimentally compared to the maximum theoretical diversity at several positions.

Structure-sequence diversification

[0093] In one embodiment, shown in Figures 9A-E, a multi-step protocol 900 can be used to generate structure-sequence diversity from the initial protein scaffold. The protocol can include: I) generation of structural ensembles 910; II) sequence design & filtering 920, III) explicit design recombination 930, and IV) conformational resampling of the interacting segment(s) 940.

[0094] In one embodiment, the Rosetta Loop-Relax functionality can be used to generate structural ensembles of the connecting segments that were built to connect the functional motif to the protein scaffold. Starting from the initial design, sets of decoys are generated separately for each of the individual segments as shown in Figure 9.

[0095] The best decoys by Rosetta energy are carried to the design stage, and for each decoy several designed models are generated. Figures 6 and 9 show positions allowed to change during the procedure. A first filter can be applied to the resulting designs to eliminate models with the same Rosetta full atom energy to reduce the number of models. A Ramachandran filter can be applied to remove designs with a worse score than the initial computational design. The complexes between the functional motif and the remaining designs are reconstituted and the interface side-chains repacked for accurate calculation of the interaction energy. Designs with interaction energy higher than 0, indicative of clashes across the interface, can be discarded. After the filtering stage designs for input motif N-terminus and C-terminus remain.

[0096] Figure 9 indicates that a set of designs selected from the single segment simulations in an all-against-all fashion can be combined. Additional modeling can be performed to ensure mutual structural compatibility between the sampled segments. The atomic coordinates from one input motif N-terminus variants are explicitly combined with the variants of other functional motif N-termini and the resulting models are then combined with other functional motif C-terminus variants, amounting to several different combinations. In some embodiments, the designs are selected based on Rosetta full-atom energy, protein core packing and the number of buried unsatisfied polar atoms. After the explicit recombination of the segments, a sequence design step can be performed where variability is allowed at the positions designed in the single segment simulations.

[0097] In some embodiments, the single segment simulations may indicate that the N-terminus of an input motif interacted with the other segments. If that is the case, a final round

of conformational sampling and design can be performed on the input motif N-terminus, to ensure the compatibility of the segments. Several designs resulting from the recombination step can be selected for re-sampling and sequence design according to Rosetta full atom-energy, or based on the conformational diversity of the rebuilt segments to enhance variety within the pool of selected designs. In general this re-sampling can be done for any segments that interact with other segments.

[0098] A subset of designs can be selected to inform an experimentally directed library. Several designs from the last ensembles generated by the re-sampling and design steps can be further filtered, using a triple criteria where the designs ranked on the top 5% by full atom energy, packing and number of buried unsatisfied polar atoms. The initial models obtained after the recombination and design steps can also be included in the design pool for library generation. The frequencies of amino acids for each designable position are computed and the computational models are visually inspected. To reduce the size of the library, some amino acids that occurred in the computational models can be discarded if their chemical composition are similar to others already included in the library, or if the filters failed to remove obvious structural flaws such as buried side chain polar atoms and under-packed models. The observed frequencies in the Rosetta models and amino acids included in the library are shown in Figures 10A and 10B.

In vitro Screening

[0099] Performing *in vitro* screening includes screening functional protein scaffolds derived from computation-guided libraries to identify clones with desired functional activity.

[00100] In some embodiments *in vitro* screening can be performed through yeast display. To overcome the limitations of the library size supported by yeast display, two partially overlapping sub-libraries can be constructed and screened sequentially as indicated in Figures 11A, 11B, 12A, and 12B. The first sub-library (Library 1) can contain computationally generated sequence diversity in connecting segments from the Primary Loop combined with design variants of Secondary segment connecting segments present in some of the models. After rounds of screening, the selected Primary Loop variants are combined with the full computationally generated diversity of the Secondary segment variants to create Library 2. This sub-library can be screened for several more rounds to isolate a clone that differs from the original computational design by several mutations distributed over the connecting segments of the Primary and Secondary segments as indicated in Figures 6 and 13.

[00101] In one example, a clone, recombinant 2bodx_42, bound b12 with a K_D of 166 nM, an improvement by a factor of >1800 over the original computational design (Table 1).

[00102] In another example, introducing the A118V mutation previously identified by random mutagenesis further increased the b12 affinity, as the resulting variant (2bodx_43) had a K_D of 33 nM (Table 1, Fig. 2, Fig. S5). This was very close to the affinity of gp120 for b12 (K_D = 1-20 nM (22, 29, 30). In another example, introducing the D114R mutation on 2bodx_43 resulted in loss of detectable b12 binding (Fig. 14), demonstrating that the high affinity was specific to the grafted epitope. Further, 2bodx_43 was thermally stable (T_m = 75 °C) and monomeric in solution (Fig. 15).

Polypeptides of the Invention, compositions thereof, and uses

[00103] In another aspect, the present invention provides isolated polypeptides, comprising or consisting of an amino acid sequence according to the following:

SEQ ID NO:1:

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVV(N/D)(P/V)(D/G)P(G/T/V)GDM(D/S)(N/Q)G(F/E/P)(E/A)(E/A)GK
QWIDEFAAGLKNRPAYIIVDP(G/L)(G/Y)SGGDPEI(A/V/L)(E/Q)(A/E)(A/W)(W/L)R(F/
T/M/S)(A/V)AYAGKALKAGSSQARIYFDAGHSAWHSPA(Q/K)(M/W)A(A/P/R)ALQRA
DISNSAHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNES
DPSGRAIGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA

[00104] The inventors have designed the polypeptides of the invention by grafting a discontinuous HIV gp120 epitope, targeted by the broadly neutralizing antibody b12, onto an unrelated scaffold. b12 binds to a conserved epitope within the CD4-binding site (CD4bs) of gp120. The CD4b is a major antibody target in HIV infection. Thus, reagents that bind b12 but not CD4b-directed non-neutralizing antibodies, such as the polypeptides of the present invention, are of great value, for example, as HIV vaccines.

[00105] Parentheses represent variable positions in the polypeptide, with the recited amino acid residues as alternatives in these positions.

[00106] In one preferred embodiment, the polypeptides comprise or consist of an amino acid sequence according to the following:

SEQ ID NO:2:

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGP(T/V)GDMSQG(E/P)(E/A)(E/A)GKQWIDEFAAGLKNRPA
YIIVDPLYSGGDPEI(V/L)QEWLR(T/M/S)VAYAGKALKAGSSQARIYFDAGHSAWHS
PA(Q/K)(M/W)A(A/P/R)ALQRADISNSAHGIATNTSNYRWTADDEVAYAKAVLSAIGNP
SLRAVIDTSRNGNGPAGNESCDPSGRAIGTPSTTNTGDPMIDAFLWIKLPGEADGCIA
GAGQFVPQAAYEMAIAA

[00107] Polypeptides according to this genus are those that are present in those polypeptides demonstrating the best range of activities, as demonstrated in the examples that follow. In a further preferred embodiment, the polypeptides comprise or consist of an amino acid sequence according to the following:

Sequence Description: SEQ ID NO. 3:

2bodx_03

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVNPDPGGMDNGFEEGKQWIDEFAAGLKNRPAYIIVDPGGSG
GDPEIAEAAWRFAAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISN
SAHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSG
RAIGTPSTTNTGDPMIDAFLWIKLPGEADGCIA GAGQFVPQAAYEMAIAA

Sequence Description: SEQ ID NO. 4:

2bodx_43

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPTGDMSQGEEAGKQWIDEFAAGLKNRPAYIIVYPLYSGG
DPEIVQEWLRTVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISNSA
HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGRA
IGTPSTTNTGDPMIDAFLWIKLPGEADGCIA GAGQFVPQAAYEMAIAA

Sequence Description: SEQ ID NO. 5:

2bodx_45

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPTGDMSQGP EAGKQWIDEFAAGLKNRPAYIIVYPLYSGG
DPEIVQEWLRMVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISNS
AHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGR
AIGTPSTTNTGDPMIDAFLWIKLPGEADGCIA GAGQFVPQAAYEMAIAA

Sequence Description: SEQ ID NO. 6:

2bidx_42

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPTGDMSQGEEAGKQWIDEFAAGLKNRPAYIIVYPLYSGG
DPEIAQEWLRTVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISNSA
HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGRA
IGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA

Sequence Description: SEQ ID NO. 7:

2bidx_44

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPGVDMSQGEEAGKQWIDEFAAGLKNRPAYIIVYPLYSG
GDPEIVQEWLRSVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISNS
AHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGR
AIGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA

SEQ ID NO. 8

b12_2bidx_043_V118L_PC

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPTGDMSQGEEAGKQWIDEFAAGLKNRPAYIIVYPLYSGG
DPEILQEWLRTVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISNSA
HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGRA
IGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAAG

SEQ ID NO. 9

b12_2bidx_043_A158R_PC

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPTGDMSQGEEAGKQWIDEFAAGLKNRPAYIIVYPLYSGG
DPEIVQEWLRTVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMA**R**ALQRADISNSA
HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGRA
IGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAAG

SEQ ID NO. 10

b12_2bodx_043_Q155K_PC

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPTGDMSQGEEAGKQWIDEFAAGLKNRPAYIIVYPLYSGG
DPEIVQEWLRTVAYAGKALKAGSSQARIYFDAGHSAWHSPA**K**MAAALQRADISNSA
HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGRA
IGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAAG

SEQ ID NO. 11

b12_2bodx_043_Q155K_M156W_A158R_PC

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDVGPTGDMSQGEEAGKQWIDEFAAGLKNRPAYIIVYPLYSGG
DPEIVQEWLRTVAYAGKALKAGSSQARIYFDAGHSAWHSPA**KWAR**ALQRADISNS
AHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGR
AIGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAAG

Sequence Description: SEQ ID NO. 3:

2bodx_03

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVNPDPGGDMNGFEEGKQWIDEFAAGLKNRPAYIIVDPGGSG
GDPEIAEAAWRFAAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISN
SAHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSG
RAIGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA

SEQ ID NO. 12

b12_2bodx_060_PC

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDAPDMGDMQTSEGVGKQWIDEFAAGLKNRPAYIIVECPLSG
GDPEI(A/V)ALLSIHCAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADIS
NSAHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPS
GRAIGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA

2bodbx_060 was designed using a control, "random" library. The same length connecting segments were employed as in the original library for 2bodbx, but in this case each position in the connecting segments was allowed to be any of the 20 amino acids. The library was screened on yeast using the same strategy as described for the computation-guided library (library 1 followed by library 2). The library screening converged on the clone 2bodbx_060 with the sequence above. The A118V point mutation is a mutation that was recovered by random mutagenesis in the original library screening of 2bodbx.

SEQ ID NO. 13

b12_2bodbx_060_A118V_PC

DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWHNQHNPGQITGQVDAL
MSAAQAAGKIPILVVDAPDMGDMQTSEGVGKQWIDEFAAGLKNRPAYIIVECPLSG
GDPEIVALLSIHCAYAGKALKAGSSQARIYFDAGHSAWHSPAQMAAALQRADISNS
AHGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNESCDPSGR
AIGTPSTTNTGDPMIDAFLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA

[00108] In a further embodiment, the polypeptide includes any resurfaced version of the listed sequences, referring to resurfacing as described in Correia et al J. Mol Biol 2011 or any related application of the concept of resurfacing.

[00109] In a further embodiment, the polypeptide includes any variant of the listed sequences obtained by adding one or more disulfide bonds.

[00110] 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 contain any suitable linker, etc. for use in any desired application, such as a peptide tag to facilitate polypeptide purification, or a T-help epitope to enhance the desired immune response. For example, the polypeptides may include C-terminal "His" tags to facilitate purification.

[00111] 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.

[00112] In a further embodiment, the polypeptides of any embodiment of the invention may further comprise a tag, such as a detectable moiety or therapeutic agent. The tag(s) can be linked to the polypeptide through covalent bonding, including, but not limited to, disulfide bonding, hydrogen bonding, electrostatic bonding, recombinant fusion and conformational bonding. Alternatively, the tag(s) can be linked to the polypeptide by means of one or more linking compounds. Techniques for conjugating tags to polypeptides are well known to the skilled artisan. Polypeptides comprising a detectable tag can be used, for example, as probes to isolate B cells that are specific for the epitope present in the polypeptide. However, they may also be used for other detection and/or analytical purposes. Any suitable detection tag can be used, including but not limited to enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, radioactive materials, positron emitting metals, and nonradioactive paramagnetic metal ions. The tag used will depend on the specific detection/analysis techniques and/or methods used such as flow cytometric detection, scanning laser cytometric detection, fluorescent immunoassays, enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), bioassays (e.g., neutralization assays), Western blotting applications, etc. When the polypeptides of the invention are used for flow cytometric detections, scanning laser cytometric detections, or fluorescent immunoassays, the tag may comprise, for example, a fluorophore. A wide variety of fluorophores useful for fluorescently labeling the polypeptides of the invention are known to the skilled artisan. When the polypeptides are used for in vivo diagnostic use, the tag can comprise, for example, magnetic resonance imaging (MRI) contrast agents, such as gadolinium diethylenetriaminepentaacetic acid, to ultrasound contrast agents or to X-ray contrast agents, or by radioisotopic labeling.

[00113] The polypeptides of the invention can also comprise a tag, such as a linker (including but not limited to an amino acid linker such as cysteine or lysine), for binding to a particle, such as a virus-like particle. As another example, the polypeptides of the invention can usefully be attached to the surface of a microtiter plate for ELISA. The polypeptides of the invention can be fused to marker sequences to facilitate purification, as described in the examples that follow. Examples include, but are not limited to, the hexa-histidine tag, the myc tag the flag tag, or leader sequences for protein expression. By way of non-limiting example, the polypeptides of the present invention can be linked to a

SLYLLPTAAAGLLLLAAQPAMAM (SEQ ID NO:14) tag (the pelB leader sequence, which may include further linker residues) that was used to express the polypeptides in the *E. coli* periplasm; it was not present in the final polypeptide products because it was cleaved during synthesis in *E. coli*. Any other suitable linker sequence could be used, including but not limited to the ompA leader. The polypeptides may further comprise an N-terminal Met residue for synthesis.

[00114] In another embodiment, a plurality of the polypeptides may be complexed to a dendrimer. Dendrimers are three dimensional, highly ordered oligomeric and/or polymeric compounds typically formed on a core molecule or designated initiator by reiterative reaction sequences adding the oligomers and/or polymers and providing an outer surface. Suitable dendrimers include, but are not limited to, "starburst" dendrimers and various dendrimer polycations. Methods for the preparation and use of dendrimers are well known to those of skill in the art.

[00115] In another embodiment, the polypeptides may be fused (via recombinant or chemical means) via their N-terminus, C-terminus, or both N- and C-termini, to an oligomerization domain. Any suitable oligomerization domain can be used. In one non-limiting embodiment, the polypeptides are fused to GCN4 variants that form trimers (hence trimers or hexamers of the fused polypeptide could be displayed). In another non-limiting embodiment, the polypeptides are fused to a fibrin foldon domain that forms trimers. In other non-limiting embodiments, the oligomerization domain could be any protein that assembles into particles, including but not limited to particles made from a (non-viral) lumazine synthase protein and particles made from (non-viral) ferritin or ferritin-like proteins.

[00116] In another embodiment, the polypeptides may be chemically conjugated to liposomes. In one non-limiting embodiment, the liposomes contain a fraction of PEGylated lipid in which the PEG groups are functionalized to carry a reactive group, and the polypeptide is chemically linked to the reactive group on the PEG. In another non-limiting embodiment, additional immune-stimulating compounds are included within the liposomes, either within the lipid layers or within the interior. In another non-limiting embodiment, specific cell-targeting molecules are included on the surface of the liposome, including but not limited to molecules that bind to proteins on the surface of dendritic cells.

[00117] In another embodiment, a plurality (ie: 2 or more; preferably at least 5, 10, 15, 20, 25, 50, 75, 90, or more copies) of the polypeptides may be present in a virus-like particle

(VLP), to further enhance presentation of the polypeptide to the immune system. As used herein, a "virus-like particle" refers to a structure that in at least one attribute resembles a virus but which has not been demonstrated to be infectious. Virus-like particles in accordance with the invention do not carry genetic information encoding for the proteins of the virus-like particles. In general, virus-like particles lack a viral genome and, therefore, are noninfectious. In addition, virus-like particles can often be produced in large quantities by heterologous expression and can be easily purified. In a preferred embodiment, the VLP comprises viral proteins that may undergo spontaneous self-assembly, including but not limited to recombinant proteins of adeno associated viruses, rotavirus, recombinant proteins of norwalkvirus, recombinant proteins of alphavirus, recombinant proteins of foot and mouth disease virus, recombinant proteins of retrovirus, recombinant proteins of hepatitis B virus, recombinant proteins of tobacco mosaic virus, recombinant proteins of flock house virus, and recombinant proteins of human papillomavirus, and Qbeta bacteriophage particles. In one preferred embodiment, the viral proteins comprise hepatitis B core antigen particles. In another embodiment, the VLPs are from lipid-enveloped viruses and include lipid as well as any suitable viral protein, including but not limited to proteins from chikungunya virus, or hepatitis B surface antigen proteins. Methods for producing and characterizing recombinantly produced VLPs have been described for VLPs from several viruses, as reviewed in US 20110236408; see also US 7,229,624. As described in the examples that follow, immunization in the context of a VLP with approximately 75 copies of the FFL_001 polypeptide (SEQ ID NO:4) conjugated onto Hepatitis B (HepB) core antigen particles results in an increased immune response to the polypeptide.

[00118] The VLPs of the invention can be used as vaccines or antigenic formulations for treating or limiting HIV infection, as discussed herein. In some embodiments, the VLPs may further comprise other scaffolds presenting other epitopes appropriate antigens. In other embodiments, the VLP may further comprise scaffolds presenting epitopes from additional HIV proteins, such as gag or pol.

[00119] In another embodiment, the polypeptides may be present on a non-natural core particle, such as a synthetic polymer, a lipid micelle or a metal. Such core particles can be used for organizing a plurality of polypeptides of the invention for delivery to a subject, resulting in an enhanced immune response. By way of example, synthetic polymer or metal core particles are described in U.S. Pat. No. 5,770,380, which discloses the use of a calixarene organic scaffold to which is attached a plurality of peptide loops in the creation of

an 'antibody mimic', and U.S. Pat. No. 5,334,394 describes nanocrystalline particles used as a viral decoy that are composed of a wide variety of inorganic materials, including metals or ceramics. Preferred metals in this embodiment include chromium, rubidium, iron, zinc, selenium, nickel, gold, silver, platinum. Preferred ceramic materials in this embodiment include silicon dioxide, titanium dioxide, aluminum oxide, ruthenium oxide and tin oxide. The core particles of this embodiment may be made from organic materials including carbon (diamond). Preferred polymers include polystyrene, nylon and nitrocellulose. For this type of nanocrystalline particle, particles made from tin oxide, titanium dioxide or carbon (diamond) are particularly preferred. A lipid micelle may be prepared by any means known in the art. See US 7,229,624 and references disclosed therein.

[00120] In another 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 polyA sequences, modified Kozak sequences, and sequences encoding epitope tags, export signals, and secretory signals, nuclear localization signals, and plasma membrane localization signals. It will be apparent to those of skill in the art, based on the teachings herein, what nucleic acid sequences will encode the polypeptides of the invention.

[00121] In a further 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, HIV, actin, EF) or inducible (driven by any of a number of inducible promoters including, but not limited to, tetracycline, ecdysone, steroid-responsive). The construction of expression vectors for use in transfecting prokaryotic cells is also well known in the art, and thus can be accomplished via standard techniques. (See, for example, Sambrook, Fritsch, and Maniatis, in: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989; *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, TX). The expression vector must be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. In a preferred embodiment, the expression vector comprises a plasmid. However, the invention is intended to include other expression vectors that serve equivalent functions, such as viral vectors.

[00122] In another aspect, the present invention provides host cells that have been transfected with the recombinant expression vectors disclosed herein, wherein the host cells can be either prokaryotic or eukaryotic. The cells can be transiently or stably transfected. Such transfection of expression vectors into prokaryotic and eukaryotic cells can be accomplished via any technique known in the art, including but not limited to standard bacterial transformations, calcium phosphate co-precipitation, electroporation, or liposome mediated-, DEAE dextran mediated-, polycationic mediated-, or viral mediated transfection. (See, for example, *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press; *Culture of Animal Cells: A Manual of Basic Technique*, 2nd Ed. (R.I. Freshney, 1987, Liss, Inc. New York, NY). A method of producing a polypeptide according to the invention is an additional part of the invention. The method comprises the steps of (a) culturing a host cell 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 medium. Methods to recover polypeptide from cell free extracts or culture medium are well known to the man skilled in the art.

[00123] In a still further aspect, the present invention provides pharmaceutical compositions (such as a vaccine), comprising one or more polypeptides, VLPs, nucleic acids, recombinant expression vectors, or host cells of the invention and a pharmaceutically

acceptable carrier. The pharmaceutical compositions of the invention can be used, for example, in the methods of the invention described below. The pharmaceutical composition may comprise in addition to the polypeptide of the invention (a) a lyoprotectant; (b) a surfactant; (c) a bulking agent; (d) a tonicity adjusting agent; (e) a stabilizer; (f) a preservative and/or (g) a buffer.

[00124] In some embodiments, the buffer in the pharmaceutical composition is a Tris buffer, a histidine buffer, a phosphate buffer, a citrate buffer or an acetate buffer. The pharmaceutical composition may also include a lyoprotectant, e.g. sucrose, sorbitol or trehalose. In certain embodiments, the pharmaceutical composition includes a preservative e.g. benzalkonium chloride, benzethonium, chlorohexidine, phenol, m-cresol, benzyl alcohol, methylparaben, propylparaben, chlorobutanol, o-cresol, p-cresol, chlorocresol, phenylmercuric nitrate, thimerosal, benzoic acid, and various mixtures thereof. In other embodiments, the pharmaceutical composition includes a bulking agent, like glycine. In yet other embodiments, the pharmaceutical composition includes a surfactant e.g., polysorbate-20, polysorbate-40, polysorbate- 60, polysorbate-65, polysorbate-80 polysorbate-85, poloxamer-188, sorbitan monolaurate, sorbitan monopalmitate, sorbitan monostearate, sorbitan monooleate, sorbitan trilaurate, sorbitan tristearate, sorbitan trioleate, or a combination thereof. The pharmaceutical composition may also include a tonicity adjusting agent, e.g., a compound that renders the formulation substantially isotonic or isoosmotic with human blood. Exemplary tonicity adjusting agents include sucrose, sorbitol, glycine, methionine, mannitol, dextrose, inositol, sodium chloride, arginine and arginine hydrochloride. In other embodiments, the pharmaceutical composition additionally includes a stabilizer, e.g., a molecule which, when combined with a protein of interest substantially prevents or reduces chemical and/or physical instability of the protein of interest in lyophilized or liquid form. Exemplary stabilizers include sucrose, sorbitol, glycine, inositol, sodium chloride, methionine, arginine, and arginine hydrochloride.

[00125] The polypeptides may be the sole active agent in the pharmaceutical composition, or the composition may further comprise one or more other agents suitable for an intended use, including but not limited to adjuvants to stimulate the immune system generally and improve immune responses overall. Any suitable adjuvant can be used. The term "adjuvant" refers to a compound or mixture that enhances the immune response to an antigen. Exemplary adjuvants include, but are not limited to, Adju-Phos, AdjuverTM, albumin-heparin microparticles, Algal Glucan, Algammulin, Alum, Antigen Formulation,

AS-2 adjuvant, autologous dendritic cells, autologous PBMC, AvridineTM, B7-2, BAK, BAY R1005, Bupivacaine, Bupivacaine-HCl, BWZL, Calcitriol, Calcium Phosphate Gel, CCR5 peptides, CFA, Cholera holotoxin (CT) and Cholera toxin B subunit (CTB), Cholera toxin A1-subunit-Protein A D-fragment fusion protein, CpG, CRL1005, Cytokine-containing Liposomes, D-Murapalmitine, DDA, DHEA, Diphtheria toxoid, DL-PGL, DMPC, DMPG, DOC/Alum Complex, Fowlpox, Freund's Complete Adjuvant, Gamma Inulin, Gerbu Adjuvant, GM-CSF, GMDP, hGM-CSF, hIL-12 (N222L), hTNF-alpha, IFA, IFN-gamma in pcDNA3, IL-12 DNA, IL-12 plasmid, IL-12/GMCSF plasmid (Sykes), IL-2 in pcDNA3, IL-2/Ig plasmid, IL-2/Ig protein, IL-4, IL-4 in pcDNA3, Imiquimod, ImmTherTM, Immunoliposomes Containing Antibodies to Costimulatory Molecules, Interferon-gamma, Interleukin-1 beta, Interleukin-12, Interleukin-2, Interleukin-7, ISCOM(s)TM, Iscoprep 7.0.3TM, Keyhole Limpet Hemocyanin, Lipid-based Adjuvant, Liposomes, Loxoribine, LT(R192G), LT-OA or LT Oral Adjuvant, LT-R192G, LTK63, LTK72, MF59, MONTANIDE ISA 51, MONTANIDE ISA 720, MPL.TM., MPL-SE, MTP-PE, MTP-PE Liposomes, Murametide, Murapalmitine, NAGO, nCT native Cholera Toxin, Non-Ionic Surfactant Vesicles, non-toxic mutant E112K of Cholera Toxin mCT-E112K, p-Hydroxybenzoique acid methyl ester, pCIL-10, pCIL12, pCMVmCAT1, pCMVN, Peptomer-NP, Pleuran, PLG, PLGA, PGA, and PLA, Pluronic L121, PMMA, PODDSTM, Poly rA: Poly rU, Polysorbate 80, Protein Cochleates, QS-21, Quadri A saponin, Quil-A, Rehydragel HPA, Rehydragel LV, RIBI, Ribilike adjuvant system (MPL, TMD, CWS), S-28463, SAF-1, Sclavo peptide, Sendai Proteoliposomes, Sendai-containing Lipid Matrices, Span 85, Specol, Squalane 1, Squalene 2, Stearyl Tyrosine, Tetanus toxoid (TT), TheramideTM, Threonyl muramyl dipeptide (TMDP), Ty Particles, and Walter Reed Liposomes. Selection of an adjuvant depends on the subject to be vaccinated. Preferably, a pharmaceutically acceptable adjuvant is used.

[00126] The compositions may further comprise other compounds useful for treating HIV infection and/or immune suppression, including reverse transcriptase inhibitors including but not limited to 3'-azido-3'-deoxythymidine (AZT), 2',3'-dideoxycytidine (DDC) and 2',3'-dideoxyinosine (DDI), zidovudine, didanosine, zalcitabine, stavudine, and viramune; protease inhibitors such as saquinovirTM, nefinavirTM, ritonavirTM, and indinavirTM; cytokines such as G-CSF, IL-11, erythropoietin, and antibiotics.

[00127] Compositions comprising the polypeptides can be stored in any standard form, including, e.g., an aqueous solution or a lyophilized cake. Such compositions are typically

sterile when administered to cells or subjects. Sterilization of an aqueous solution is readily accomplished by filtration through a sterile filtration membrane. If the composition is stored in lyophilized form, the composition can be filtered before or after lyophilization and reconstitution.

[00128] In another aspect, the present invention provides methods for treating and/or limiting an HIV infection, comprising administering to a subject in need thereof a therapeutically effective amount of one or more polypeptides of the invention, salts thereof, conjugates thereof, VLPs thereof, or pharmaceutical compositions thereof, to treat and/or limit the HIV infection. In another embodiment, the method comprises eliciting an immune response in an individual having or at risk of an HIV infection, comprising administering to a subject in need thereof a therapeutically effective amount of one or more polypeptides of the invention, salts thereof, conjugates thereof, VLPs thereof, or pharmaceutical compositions thereof, to generate an immune response.

[00129] "Human Immunodeficiency Virus" and "HIV" includes all variants and types of HIV-1, HIV-2, and other synonymous retroviruses, such as human T-lymphotropic virus type III (HTLV-III) and lymphadenopathy associated virus (LAV-1 and LAV-2).

[00130] When the method comprises treating an HIV infection, the one or more polypeptides, VLPs, or compositions are administered to a subject that has already been infected with the HIV, and/or who is suffering from symptoms (including but not limited to acquired immune deficiency syndrome (AIDS), susceptibility to pathogenic and opportunistic organisms and infections, anemia, thrombocytopenia, and lymphopenia) indicating that the subject is likely to have been infected with the HIV. As used herein, the term "opportunistic infection" refers to infections with an organism that would not normally be pathologic in patients with intact immune systems.

[00131] As used herein, "treat" or "treating" means accomplishing one or more of the following: (a) reducing HIV titer in the subject; (b) limiting any increase of HIV titer in the subject; (c) reducing the severity of HIV symptoms; (d) limiting or preventing development of HIV symptoms after infection; (e) inhibiting worsening of HIV symptoms; (f) limiting or preventing recurrence of HIV symptoms in subjects that were previously symptomatic for HIV infection. In one embodiment method, polypeptides, VLPs, or compositions are used as "therapeutic vaccines" to ameliorate the existing infection and/or provide prophylaxis against infection with additional HIV virus.

[00132] When the method comprises limiting an HIV infection, the one or more polypeptides, VLPs, or compositions can also be administered prophylactically to a subject that is not known to be infected, but may be at risk of exposure to the HIV. As used herein, “limiting” means to limit HIV infection in subjects at risk of HIV infection. Groups at particularly high risk include individuals having unprotected sex with multiple partners, individuals having sex with HIV-infected partners, individuals in need of blood transfusions, and infants nursing from HIV-infected mothers. In this method, the polypeptides, VLPs, or compositions are used as vaccines.

[00133] As used herein, a “therapeutically effective amount” refers to an amount of the polypeptide that is effective for treating and/or limiting HIV infection. The polypeptides are typically formulated as a pharmaceutical composition, such as those disclosed above, and can be administered via any suitable route, including orally, parentally, by inhalation spray, rectally, or topically in dosage unit formulations containing conventional pharmaceutically acceptable carriers, adjuvants, and vehicles. The term parenteral as used herein includes, subcutaneous, intravenous, intra-arterial, intramuscular, intrasternal, intratendinous, intraspinal, intracranial, intrathoracic, infusion techniques or intraperitoneally. Polypeptide compositions may also be administered via microspheres, liposomes, immune-stimulating complexes (ISCOMs), or other microparticulate delivery systems or sustained release formulations introduced into suitable tissues (such as blood). Dosage regimens can be adjusted to provide the optimum desired response (e.g., a therapeutic or prophylactic response). A suitable dosage range may, for instance, be 0.1 ug/kg-100 mg/kg body weight; alternatively, it may be 0.5 ug/kg to 50 mg/kg; 1 ug/kg to 25 mg/kg, or 5 ug/kg to 10 mg/kg body weight. The polypeptides can be delivered in a single bolus, or may be administered more than once (e.g., 2, 3, 4, 5, or more times) as determined by an attending physician.

[00134] In certain embodiments, the polypeptides of the invention neutralize HIV infectivity. In various embodiments, the polypeptides of the invention prevent HIV from infecting host cells by at least 99%, at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 60%, at least 50%, at least 45%, at least 40%, at least 35%, at least 30%, at least 25%, at least 20%, or at least 10% relative to infection of host cells by HIV in the absence of the polypeptides. Neutralization can be measured using standard techniques in the art.

[00135] In another aspect, the present invention provides pharmaceutical composition, comprising (a) isolated nucleic acids, recombinant expression vectors, and/or recombinant host cells of the invention ; and (b) a pharmaceutically acceptable carrier.

[00136] In this aspect, the nucleic acids, expression vectors, and host cells of the invention can be used as polynucleotide-based immunogenic compositions, to express an encoded polypeptide in vivo, in a subject, thereby eliciting an immune response against the encoded polypeptide. Various methods are available for administering polynucleotides into animals. The selection of a suitable method for introducing a particular polynucleotide into an animal is within the level of skill in the art. Polynucleotides of the invention can also be introduced into a subject by other methods known in the art, e.g., transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (lysosome fusion), or a DNA vector transporter (see, e.g., Wu et al. (1992) J. Biol. Chem. 267:963-967).

[00137] The immune response against the polypeptides, VLPs, or compositions of the invention can be generated by one or more inoculations of a subject with an immunogenic composition of the invention. A first inoculation is termed a "primary inoculation" and subsequent immunizations are termed "booster inoculations". Booster inoculations generally enhance the immune response, and immunization regimens including at least one booster inoculation are preferred. Any polypeptide, VLP, or composition of the invention may be used for a primary or booster immunization. The adequacy of the vaccination parameters chosen, e.g., formulation, dose, regimen and the like, can be determined by taking aliquots of serum from the subject and assaying antibody titers during the course of the immunization program. Alternatively, the T cell populations can be monitored by conventional methods. In addition, the clinical condition of the subject can be monitored for the desired effect, e.g., limiting HIV infection, improvement in disease state (e.g., reduction in viral load), etc. If such monitoring indicates that vaccination is sub-optimal, the subject can be boosted with an additional dose of composition, and the vaccination parameters can be modified in a fashion expected to potentiate the immune response. Thus, for example, the dose of the polypeptide, VLP, or composition, and/or adjuvant, can be increased or the route of administration can be changed.

[00138] In a further aspect, the present invention provides methods for monitoring HIV infection or an HIV-induced disease in a subject and/or monitoring response of the subject to

immunization by an HIV vaccine, comprising contacting the polypeptides, the VLPs, or the pharmaceutical compositions of the invention with a bodily fluid from the subject and detecting HIV-binding antibodies in the bodily fluid of the subject. By "HIV-induced disease" is intended any disease caused, directly or indirectly, by HIV, including but not limited to AIDS. The method comprises contacting a polypeptide, VLP, or composition of the invention with an amount of bodily fluid (such as serum, whole blood, etc.) from the subject; and detecting HIV-binding antibodies in the bodily fluid of the subject. The detection of the HIV binding antibodies allows the HIV infection progression and/or HIV-induced disease in the subject to be monitored. In addition, the detection of HIV binding antibody also allows the response of the subject to immunization by an HIV vaccine to be monitored. In still other methods, the titer of the HIV binding antibodies is determined. Any suitable detection assay can be used, including but not limited to homogeneous and heterogeneous binding immunoassays, such as radioimmunoassays (RIA), ELISA, immunofluorescence, immunohistochemistry, FACS, BIACORE and Western blot analyses. The methods may be carried in solution, or the polypeptide(s) of the invention may be bound or attached to a carrier or substrate, e.g., microtiter plates (ex: for ELISA), membranes and beads, etc. Carriers or substrates may be made of glass, plastic (e.g., polystyrene), polysaccharides, nylon, nitrocellulose, or teflon, etc. The surface of such supports may be solid or porous and of any convenient shape. The polypeptides of the invention for use in this aspect may comprise a conjugate as disclosed above, to provide a tag useful for any detection technique suitable for a given assay.

[00139] In a still further aspect, the present invention provides methods for detecting HIV binding antibodies, comprising: (a) contacting the polypeptides, the VLPs, or the compositions of the invention with a composition comprising a candidate HIV binding antibody under conditions suitable for binding of HIV antibodies to the polypeptide, VLP, or composition; and (b) detecting HIV antibody complexes with the polypeptide, VLP, or composition. In this aspect, the methods are performed to determine if a candidate HIV binding antibody recognizes the HIV b12 epitope present in the polypeptides of the invention. Any suitable composition may be used, including but not limited to bodily fluid samples (such as serum, whole blood, etc.) from a suitable subject (such as one who has been infected with HIV), naive libraries, modified libraries, and libraries produced directly from human donors exhibiting an HIV-specific immune response. The assays are performed under conditions suitable for promoting binding of antibodies against the polypeptides; such

conditions can be determined by those of skill in the art based on the teachings herein. Any suitable detection assay can be used, including but not limited to homogeneous and heterogeneous binding immunoassays, such as radioimmunoassays (RIA), ELISA, immunofluorescence, immunohistochemistry, FACS, BIACORE and Western blot analyses. The methods may be carried in solution, or the polypeptide(s) of the invention may be bound or attached to a carrier or substrate, e.g., microtiter plates (ex: for ELISA), membranes and beads, etc. Carriers or substrates may be made of glass, plastic (e.g., polystyrene), polysaccharides, nylon, nitrocellulose, or teflon, etc. The surface of such supports may be solid or porous and of any convenient shape. The polypeptides of the invention for use in this aspect may comprise a conjugate as disclosed above, to provide a tag useful for any detection technique suitable for a given assay. In a further embodiment, the HIV b12-binding antibodies are isolated using standard procedures. In one embodiment, the methods may comprise isolation of polypeptide-specific memory B cells by fluorescence activated cell sorting (FACS) using standard techniques in the art (see, for example, *Science* DOI: 10.1126/science.1187659).

[00140] In another aspect, the present invention provides methods for producing HIV antibodies, comprising (a) administering to a subject an amount effective to generate an antibody response of the polypeptides, the VLPs, and/or the compositions of the invention; and (b) isolating antibodies produced by the subject.

[00141] The polypeptides of the invention can also be used to generate antibodies that recognize the polypeptides of the invention. The method comprises administering to a subject a polypeptide, VLP, or composition of the invention. Such antibodies can be used, for example, in HIV research. A subject employed in this embodiment is one typically employed for antibody production, including but not limited to mammals, such as, rodents, rabbits, goats, sheep, etc. The antibodies generated can be either polyclonal or monoclonal antibodies. Polyclonal antibodies are raised by injecting (e.g. subcutaneous or intramuscular injection) antigenic polypeptides into a suitable animal (e.g., a mouse or a rabbit). The antibodies are then obtained from blood samples taken from the animal. The techniques used to produce polyclonal antibodies are extensively described in the literature. Polyclonal antibodies produced by the subjects can be further purified, for example, by binding to and elution from a matrix that is bound with the polypeptide against which the antibodies were raised. Those of skill in the art will know of various standard techniques for purification and/or

concentration of polyclonal, as well as monoclonal, antibodies. Monoclonal antibodies can also be generated using techniques known in the art.

Examples

Scaffold Matching for the b12 epitope

[00142] To identify a set of scaffolds that could accommodate the gp120 epitope when bound to b12 the Fab structure, the matching stage exhaustively searched 13337 single chains from the PDB. The input motif was composed of 2 segments, which will be referred as the CD4b loop (residue 364-373 from gp120, PDBid: 2NY7) and the ODe loop (residue 472-476 from gp120, PDBid: 2NY7). A variety of matching schemes were used to select 11 candidate scaffolds from approximately 30,000 protein chains in the PDB. 6 candidate scaffolds were identified from superposition alignments (with either the CD4b loop or the ODe loop as the primary loop); 4 were identified from C2N alignments with the CD4b loop as the primary loop; in one scaffold the motif was matched using the End-point alignment with the CD4b loop as the primary loop. The designed scaffold based on the structure with the PDBid 2bod, which is the focus of this report, was selected using a C2N alignment with a primary CD4b loop. In all cases, the defined rmsd thresholds were 3 Å for the primary matches and 5 Å for the secondary matches. In addition to the built-in automated filters, a human evaluation step discarded matches with two problematic features: matches located in the protein core; perpendicular orientations between the backbones of the input motif and the accepting scaffold, a functionality to monitor the angles was implemented in a later version of the software.

Design of the 2bodx b12 epitope-scaffold

[00143] The hybrid structure of the 2bod scaffold was obtained as follows: residue 372 of the CD4b loop (gp120 residues 365-372) was superimposed onto residue 130 of 2bod and the CD4b loop replaced the native protein segment between residues 115-130; the ODe loop (gp120 residues 472-476) retained its original orientation relative to the CD4b loop and replaced residues 73 to 92 of the native 2bod. The backbone rebuilding and design were performed in the presence of the b12 antibody structure. A four residue segment was built to connect residues 114 (2bod numbering) and 365 of the CD4b loop (gp120 numbering). The ODe loop was connected to the scaffold by four and three residue connecting segments to the N and C termini respectively. To facilitate loop closure both terminal residues of the gp120 loops were allowed to move.

[00144] Purified designs were tested for b12 binding by surface plasmon resonance (SPR). One design, 2bodx_03, which had 39 mutations and 11 deletions relative to the parent protein (Fig. 6), bound to b12 weakly ($K_D \approx 300 \mu\text{M}$). The binding of 2bodx_03 to b12 was specific, because no binding was detected for a ‘dead epitope’ mutant (D114R) equivalent to the D368R mutation on gp120 that ablates b12 binding (25) (Fig. 7). In an initial attempt to optimize the b12 affinity of 2bodx_03, a random mutagenesis library was constructed and screened by yeast display (26). This led to the identification of clone 2bodx_R3 that had two mutations (S177G, A118V) from 2bodx_03 and showed one order of magnitude higher affinity for b12 ($K_D \approx 30 \mu\text{M}$) by SPR (Table 1, Fig. 2, Fig. S5). However, this interaction was still three orders of magnitude weaker than the b12 interaction with gp120.

Library screening on the surface of yeast

[00145] Genes encoding the computational designs were synthesized with optimized codon usage and subcloned into pET29b vector (Codon Devices, Genscript). For library construction, these genes were extracted and extended by PCR to add regions of homology to the pCTCON2 vector.

[00146] The random mutagenesis library was generated with the Gene Morph II Random Mutagenesis kit (Stratagene) under reaction conditions that resulted in ~ 4 mutations per gene (60 ng starting DNA template). For the directed libraries, CD4b loop or ODe loop-encoding ultramer oligos (IDT) were transformed into double stranded DNA by PCR and combined with two non-mutated DNA fragments extracted from the original 2bodx_03 gene to assemble full-length genes via homologous recombination in yeast (Figs. 11A and 11B). The sequence variants explored in the three libraries as well as the theoretical and experimental library sizes are shown in Figs. 11A, 11B, 12A, and 12B.

[00147] Libraries were transformed and screened using a yeast surface display system previously described. Briefly, *S. cerevisiae* EBY100 competent cells were transformed with $1 \mu\text{g}$ of pCTCON2 triple-cut vector (BamHI/SalI/NheI, NEB) and a 20x molar excess of library DNA using electroporation. Typical transformation efficiency ranged from 10^6 to 10^7 for a 1x transformation and $\sim 80\%$ of the recovered sequences from the naïve libraries contained full-length in-frame gene variants. The resulting yeast culture was grown at 30°C , with 250 rpm shaking in 2% glucose C-trp-ura media and passaged at least two times. Once the culture reached a density of 2×10^7 - 3×10^7 cells/mL it was transferred to 2% galactose C-trp-ura media and grown for 14-16 hours to induce protein expression on the surface of yeast.

106 to 108 yeast cells were then pelleted and washed three times with PBSA buffer (0.01 M sodium phosphate, pH 7.4, 0.137 M sodium chloride, 1 g/L bovine serum albumin). Cells were then incubated with b12 IgG on an orbital shaker for 1 hour at 4 °C. Finally, cells were pelleted, washed three times and fluorescently labeled with phycoerythrin conjugated α -hIgG (Invotrogen) and fluorescein isothiocyanate labeled α -cMyc Ab (Immunology Consultants Laboratory) at 4 °C for 0.5-1 hour.

[00148] Cells were analyzed using fluorescence activated cell sorting (BD Influx, BD Biosciences) and double positive clones were collected, expanded, induced and labeled as before for additional rounds of selection. Decreasing monoclonal antibody (mAb) b12 IgG concentrations were used for labeling subsequent selection rounds to reduce library diversity and select high affinity clones. The initial random library was screened for 5 rounds with 1 μ M b12 IgG. Library 1 was screened for 5 rounds at 1 μ M, 1 μ M, 100 nM, 10 nM and 1 nM b12 IgG respectively. Library 2 was screened for 3 rounds with 10 nM, 100 pM and 10 pM b12 IgG respectively. Library 3 was screened for 4 rounds with 1 nM, 1 nM, 100 pM and 10 pM b12 IgG respectively. Libraries 2 and 3 were rescreened with b12 Fab to avoid any selection artifacts due to the bivalence of the labeling reagent. The same clones isolated by IgG screening were predominant after selection rounds with Fab. . FACS data was analyzed using FlowJo 8.8.6 (Tree Star) as described before (10).

[00149] Individual clones were selected after each round and their DNA was isolated using the Zymoprep II kit (Zymo Research). Standard PCR was used to amplify the gene of interest from this DNA and the resulting product was sent for sequence analysis.

Protein expression, purification and characterization

[00150] DNA segments encoding scaffold constructs were synthesized with optimized codon usage and RNA structure (Codon Devices, Genscript Corp.), subcloned into pET29 (EMD Biosciences) and transformed into Arctic Express E. coli (Invitrogen). Single colonies were grown overnight at 37 °C in 10 mL Luria Broth (LB) plus Kanamycin (100 mg/ml). The starter cultures were expanded into 1 L of LB plus Kanamycin and incubated at 37 °C; when cells reached log phase, 250 μ M of IPTG was added to the cultures to induce protein expression and the cells were then incubated overnight at 12 °C. Cultures were then pelleted and resuspended in Start Buffer (160 mM Imidazole, 4 M Sodium Chloride, 160 mM Sodium Phosphate), a tablet of protease inhibitor (Novagen) was added and the cell suspension was frozen at -20 °C.

[00151] The cell suspension was thawed and 10 ml of 10X Bugbuster (Novagen), 50 μ L of Benzonase Nucleases and 1.7 μ L of rLysozyme (Novagen) were added to lyse the cells; the cell suspension was then gently tumbled in an orbital shaker for 20 minutes. Lysed cells were pelleted and the supernatant was filtered through a 0.22 μ m filter (Millipore). Supernatants were tumbled with 5 mL of Ni⁺⁺ Sepharose 6 Fast Flow (GE Healthcare) for 1 hour at 4 °C. The resin was washed 3 times with 30 mL Wash Buffer (50 mM imidazole, 500 mM Sodium Chloride and 160 mM Sodium Phosphate) and eluted with 20 mL of Elution Buffer (250 mM Imidazole, 500 mM Sodium Chloride and 20 mM Sodium Phosphate). Fractions containing the construct of interest were combined and further purified by preparative size exclusion chromatography (SEC) on Superdex 75 16/60 (GE Healthcare) at room temperature in HBS. Collected fractions were analyzed on a 4-12% SDS denaturing gel (Invitrogen) and positive fractions were combined and concentrated by ultrafiltration (Vivaspin, Bioexpress). Protein concentration was determined by measuring UV absorption signal at 280 nm (Nanodrop) and calculated from the theoretical extinction coefficient. Genes encoding different 2bodb protein variants were generated by Kunkel mutagenesis (12, 13).

[00152] The monodispersity and molecular weight of purified proteins were further assessed by HPLC (Agilent, 1200 series) coupled to an on-line static light scatter (miniDAWN TREOS, Wyatt). 120 μ L of 2-5 mg/mL protein sample was used and the collected data was analyzed with the ASTRA software (Wyatt).

[00153] Solution thermostabilities (T_m) were determined by circular dichroism (CD) on an Aviv 62A DS spectrometer. Far-UV wavelength scans (190-260 nm) of 26 μ M protein were collected in a 1mm path length cuvette. Temperature-induced protein denaturation was followed by change in ellipticity at 219 nm. Experiments were carried over a temperature range from 1-99 °C, with 2 °C increments every 3 minutes, and the resulting data was converted to mean residue ellipticity and fitted to a two-state model.

Surface Plasmon Resonance

[00154] All experiments were carried out on a Biacore 2000 (GE Healthcare) at 25 °C with HBSEP (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA and 0.005% (v/v) Surfactant P20) (GE Healthcare) as running buffer. For binding analysis, 200-500 response units (RUs) of b12 IgG were captured on a CM5 sensor chip containing 8000-9000 RUs of amine-linked mouse anti-human IgG (Human Antibody Capture kit, GE Healthcare). Samples of different protein concentrations were injected in duplicates over this surface at a

flow rate of 50-100 $\mu\text{l}/\text{min}$. If necessary, surface regeneration was performed with two 60 seconds injections of 3 M MgCl_2 at a flow rate of 10 $\mu\text{l}/\text{min}$. One flow cell contained anti-human IgG only and its interaction with the analyte was used as reference.

[00155] Data preparation and analysis were performed using Scrubber 2.0 (BioLogic Software). For kinetic analysis, biosensor data were globally fit to a mass transport limited simple bimolecular binding model:



[00156]

[00157] where A_0 represents injected analyte.

[00158] For equilibrium analysis each data set was fitted to a single site interaction model:

$$R_{eq} = \frac{R_{max} * C_A}{C_A + K_D}$$

[00159]

[00160] where R_{eq} is the response value at equilibrium, C_A is the concentration of the analyte, R_{max} is the maximum response obtained when all binding sites are occupied by the analyte and K_D is the dissociation constant.

[00161] The interaction of 2bodbx_43 with other CD4 binding site antibodies was tested by capturing ~ 400 RUs of 10 such antibodies (b12, m18, m14, 15E, b6, b13, F105, VRC01, VRC03, CD4-IgG) on two different flow cells of a CM5 chip with amine-coupled anti-human-IgG. 2bodbx_43 at 1.2 μM was injected over this surface and after the signal returned to baseline, gp120 (96 nM) was injected to confirm the reactivity of the captured antibodies.

Crystallography

2bodbx_43-b12 Fab complex

[00162] b12 Fab was produced by papain digestion of b12 IgG using the Fab Preparation Kit (Thermo Scientific). Stable complexes of 2bodbx_43 and b12 Fab were isolated by size exclusion chromatography on a SuperDex 200 16/60 column (GE Healthcare). Diffraction quality complex crystals were grown at ambient temperature by vapor diffusion of hanging drops composed of 1.4:2.0 volume ratios of protein solution (10.3 mg/mL) and reservoir solution (18-22% w:v PEG-3350, 0.14-0.22 M KNO_3). Crystals were transferred to a

cryoprotectant solution composed of reservoir solution plus 15% v:v glycerol prior to looping and flash-cooling in a 100°K nitrogen cryostat. X-ray diffraction data were collected at beamline 5.0.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) and processed with HKL-2000 (14). Initial phases were determined by molecular replacement ($z = 1$) with PHASER (15) by sequential placement of the four Fab b12 immunoglobulin domains (VL, CL, VH and CH1; PDB accession code 2NY7) and the scaffold template (PDB accession code 2BOD); residues expected to be in proximity to the engineered epitope region or the Fab hinge regions were deleted from the search models. Iterative model building was performed with COOT and refinement with REFMAC5; B-factors were modeled by TLS refinement (two TLS groups per chain) with non-crystallographic symmetry restraints imposed. Data collection and refinement statistics are presented in Table S1. Structure validation was carried out with Procheck, the MolProbity server, and the RCSB ADIT validation server.

[00163] In *computation-guided library design*, we used a structure-sequence diversification protocol (to devise relatively small libraries based on more complete sampling of low energy structures and sequences in the connecting segments. For each connecting segment, 20000 backbone conformations were separately generated and subjected to sequence design while keeping the rest of the 2bodx_03 structure fixed. Several low energy models for each segment were exhaustively recombined in silico and subjected to further sequence design to identify 2bodx models with optimal structures and sequences in all connecting segments. Following a final round of conformational resampling and design, the best 45 models by several Rosetta metrics were used to generate sequence profiles to identify the amino acids that occurred at each of the 21 positions in the connecting segments. The diversity was reduced by eliminating residues that occurred at low frequency, that were similar in size and chemical nature to more frequent residues, or that were judged likely to bury a polar side chain. The final library allowed mutations at 21 positions and had a theoretical size of 10^{12} .

[00164] For *in vitro screening*, we employed yeast display. To overcome the limitations of the library size supported by yeast display (10^7), two partially overlapping sub-libraries were constructed and screened sequentially. The first sub-library (library 1) contained all (4×10^6) of the computationally designed ODe loop connecting segments combined with 8 design variants of the CD4b loop connecting segments present in 23 of the 45 models. After three rounds of screening, the selected ODe loop variants (from at least 18 different clones)

were combined with all (2×10^5) of the computationally designed CD4b loop variants to create library 2. This sub-library was screened for three rounds to isolate clone 2bodx_42 that differed from 2bodx_03 by 17 mutations (fig. S11). Recombinant 2bodx_42 bound b12 with a K_D of 166 nM, an improvement by a factor of >1800 over 2bodx_03 (Table 1). Introducing the A118V mutation from 2bodx_R3 further increased b12 affinity, as the resulting variant (2bodx_43) bound b12 with a K_D of 33 nM (Table 1), within a factor of 2 of the b12-gp120 affinity. Introducing the D114R mutation on 2bodx_43 resulted in loss of detectable b12 binding, demonstrating that the binding was specific to the epitope. Further, 2bodx_43 was thermally stable ($T_m = 75^\circ\text{C}$) and monomeric in solution.

[00165] To assess whether the b12 affinity could be improved further and to evaluate if the computation-guided libraries restricted the sequence space effectively, we screened a third library based on 2bodx_43 with expanded sampling at 7 positions. The highest affinity clone isolated (2bodx_45) differed from 2bodx_43 by two mutations and bound b12 with a K_D of 10 nM, a factor of 3 better than 2bodx_43 and as tightly as gp120 (Table 1). Another high affinity variant selected from this library (2bodx_44, $K_D=19$ nM) was used to investigate the b12 binding contributions of library-selected mutations. We measured the b12 binding of 2bodx_44 constructs in which the “evolved” residues were individually reverted to their 2bodx_03 identity. Only 6 out of 16 reversions reduced the b12 affinity of 2bodx_44 by a factor of three or more, and a 2bodx_03 variant that contained 9 of the 2bodx_44 mutations had only micromolar affinity for b12 ($K_D = 1.5 \mu\text{M}$). Thus the selected mutations made synergistic contributions to the high b12 affinity of 2bodx constructs.

[00166] To evaluate the degree to which the 2bodx-b12 interaction recapitulated the gp120-b12 interaction, a 2.07 Å resolution crystal structure was solved for 2bodx_43 complexed with b12. Comparison with the gp120-b12 complex revealed a high degree of mimicry – superposition of the epitope and paratope of both complexes gave an overall backbone root mean square deviation (rmsd) of 0.71 Å. Consistent with good backbone mimicry, important interactions involving b12 heavy chain residues Y53, Y98, W100, N100g, and Y100h were recapitulated in the 2bodx_43-b12 complex. The total buried areas in the complexes were also similar, except for a small additional area on the scaffold outside the epitope.

[00167] The CD4bs is a major antibody target in HIV infection. Reagents are desired that bind b12 but not CD4bs-directed non-neutralizing antibodies such as b13 that engages

gp120 similarly to b12. Of 8 CD4bs-directed antibodies tested, 2bodx_43 bound tightly to b12 only. Additional SPR analyses showed that 2bodx_43 binds more tightly to b12 than to b13 by a factor of >10,000.

2bodx variant	Origin	b12 interaction parameters (SPR)			
		k_{on} ($\text{M}^{-1} \times \text{s}^{-1}$)	k_{off} (s^{-1})	K_{D} (kinetic) (nM)	K_{D} (equilibrium) (nM)
03	Initial Design	-	-	-	$> 300 \times 10^3$
Y3	RL	-	-	-	$\approx 30 \times 10^3$
42	L1+ L2	1.3×10^6	2.3×10^{-1}	177	166.6
43	L1+L2+RL	3.0×10^6	1.0×10^{-1}	33.3	33.5
44	L1+L2+RL+L3	1.9×10^6	3.6×10^{-2}	18.9	19.5
45	L1+L2+RL+L3	3.8×10^6	3.9×10^{-2}	10.3	10.3

Table 1. Affinity and kinetics of the interaction between recombinant 2bodx variants and b12 mAb. For all the reported values, the standard error is $\leq \pm 7$ of the last significant figure. RL: random library; L1: Library 1; L2: Library 2; L3: Library 3.

Example Computing Environment

[00168] Figure 19 is a block diagram of an example computing network. Some or all of the above-mentioned techniques, including but not limited to all of the above-mentioned *in silico* techniques, can be performed by a computing device. For example, Figure 19 shows protein designer 1902 configured to communicate, via network 1906, with client devices 1904a, 1904b. Protein designer 1902 can be a computing device, such as described in detail below, configured to perform part or all of the Multigraft Match and/or Multigraft Design algorithms described herein.

[00169] Network 1906 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 1906 may also correspond to a combination of one or more LANs, WANs, corporate intranets, and/or the public Internet.

[00170] Although Figure 19 only shows two client devices 1904a, 1904b, distributed application architectures may serve tens, hundreds, or thousands of client devices. Moreover, client devices 1904a and 1904b (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 1904a and 1904b can be dedicated to protein research. In other

embodiments, client devices 1904a and 1904b can be used as general purpose computers that are configured to perform a number of tasks and need not be dedicated to protein research. In still other embodiments, part or all of the functionality of protein designer 1902 and/or protein data base 1910 can be incorporated in a client device, such as client device 1904a and/or 1904b. In even other embodiments, the functionality of protein data base 1910 can be incorporated into protein designer 1902. In particular embodiments, protein data base 1910 can perform some or all of the functions described above for the protein structure database and/or the PDB. While Figure 19 shows protein designer 1902 and protein data base 1910 connected via network 1906, in some embodiments, protein designer 1902 and protein data base 1910 can be directly connected.

Computing Device Architecture

[00171] Figure 20A is a block diagram of an example computing device (e.g., system) in accordance with an example embodiment. In particular, computing device 2000 shown in Figure 20A can be configured to: (i) perform one or more functions of client device 1904a, 1904b, network 1906, and/or protein database 1910, (ii) carry out one or more of the Multigraph Match and Multigraph Design algorithms, (iii) carry out part or all of any other herein-described methods, such as but not limited to method 100. Computing device 2000 may include a user interface module 2001, a network-communication interface module 2002, one or more processors 2003, and data storage 2004, all of which may be linked together via a system bus, network, or other connection mechanism 2005.

[00172] User interface module 2001 can be operable to send data to and/or receive data from external user input/output devices. For example, user interface module 2001 can be configured to 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 2001 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 (DLP) technology, printers, light bulbs, and/or other similar devices, either now known or later developed. User interface module 2001 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.

[00173] Network-communications interface module 2002 can include one or more wireless interfaces 2007 and/or one or more wireline interfaces 2008 that are configurable to communicate via a network, such as network 1906 shown in Figure 19. Wireless interfaces 2007 can include one or more wireless transmitters, receivers, and/or transceivers, such as a 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 2008 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 more wires, a coaxial cable, a fiber-optic link, or a similar physical connection to a wireline network.

[00174] In some embodiments, network communications interface module 2002 can be configured to provide reliable, secured, and/or authenticated communications. For each communication described herein, information for ensuring reliable communications (*i.e.*, 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 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.

[00175] Processors 2003 can include one or more general purpose processors and/or one or more special purpose processors (*e.g.*, digital signal processors, application specific integrated circuits, etc.). Processors 2003 can be configured to execute computer-readable program instructions 2006 contained in data storage 2004 and/or other instructions as described herein. Data storage 2004 can include one or more computer-readable storage media that can be read and/or accessed by at least one of processors 2003. The one or more computer-readable storage media can include volatile and/or non-volatile storage 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 2003. In some embodiments, data storage 2004 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 2004 can be implemented using two or more physical devices.

[00176] Data storage 2004 can include computer-readable program instructions 2006 and perhaps additional data. For example, in some embodiments, data storage 2004 can store part or all of a protein database, such as protein database 1910. In some embodiments, data storage 2004 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.

[00177] Figure 20B depicts a network 1906 of computing clusters 2009a, 2009b, 2009c arranged as a cloud-based server system in accordance with an example embodiment. Protein database 1910 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 database 1910 can be a single computing device residing in a single computing center. In other embodiments, protein database 1910 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.

[00178] In some embodiments, data and services for protein database 1910 can be encoded as computer readable information stored in tangible computer readable media (or computer readable storage media) and accessible by client devices 1904a and 1904b, and/or other computing devices. In some embodiments, data of protein database 1910 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.

[00179] Figure 20B depicts a cloud-based server system in accordance with an example embodiment. In Figure 20B, the functions of protein designer 1902 and/or protein database 1910 can be distributed among three computing clusters 2009a, 2009b, and 2008c. Computing cluster 2009a can include one or more computing devices 2000a, cluster storage arrays 2010a, and cluster routers 2011a connected by a local cluster network 2012a. Similarly, computing cluster 2009b can include one or more computing devices 2000b, cluster storage arrays 2010b, and cluster routers 2011b connected by a local cluster network 2012b. Likewise, computing cluster 2009c can include one or more computing devices 2000c, cluster storage arrays 2010c, and cluster routers 2011c connected by a local cluster network 2012c.

[00180] In some embodiments, each of the computing clusters 2009a, 2009b, and 2009c 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.

[00181] In computing cluster 2009a, for example, computing devices 2000a can be configured to perform various computing tasks of protein designer 1902 and/or protein database 1910. In one embodiment, the various functionalities of protein designer 1902 and/or protein database 1910 can be distributed among one or more of computing devices 2000a, 2000b, and 2000c. Computing devices 2000b and 2000c in computing clusters 2009b and 2009c can be configured similarly to computing devices 2000a in computing cluster 2009a. On the other hand, in some embodiments, computing devices 2000a, 2000b, and 2000c can be configured to perform different functions.

[00182] In some embodiments, computing tasks and stored data associated with protein designer 1902 and/or protein database 1910 can be distributed across computing devices 2000a, 2000b, and 2000c based at least in part on the processing requirements of protein designer 1902 and/or protein database 1910, the processing capabilities of computing devices 2000a, 2000b, and 2000c, 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.

[00183] The cluster storage arrays 2010a, 2010b, and 2010c of the computing clusters 2009a, 2009b, and 2009c can be data storage arrays that include disk array controllers configured to 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 one or more computing devices from accessing one or more cluster storage arrays.

[00184] Similar to the manner in which the functions of protein designer 1902 and/or protein database 1910 can be distributed across computing devices 2000a, 2000b, and 2000c of computing clusters 2009a, 2009b, and 2009c, various active portions and/or backup portions of these components can be distributed across cluster storage arrays 2010a, 2010b, and 2010c. For example, some cluster storage arrays can be configured to store the data of protein designer 1902, while other cluster storage arrays can store data of protein database 1910. Additionally, some cluster storage arrays can be configured to store backup versions of data stored in other cluster storage arrays.

[00185] The cluster routers 2011a, 2011b, and 2011c in computing clusters 2009a, 2009b, and 2009c can include networking equipment configured to provide internal and external communications for the computing clusters. For example, the cluster routers 2011a in computing cluster 2009a can include one or more internet switching and routing devices configured to provide (i) local area network communications between the computing devices 2000a and the cluster storage arrays 2001a via the local cluster network 2012a, and (ii) wide area network communications between the computing cluster 2009a and the computing clusters 2009b and 2009c via the wide area network connection 2013a to network 1906. Cluster routers 2011b and 2011c can include network equipment similar to the cluster routers 2011a, and cluster routers 2011b and 2011c can perform similar networking functions for computing clusters 2009b and 2009b that cluster routers 2011a perform for computing cluster 2009a.

[00186] In some embodiments, the configuration of the cluster routers 2011a, 2011b, and 2011c 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 equipment in the cluster routers 2011a, 2011b, and 2011c, the latency and throughput of local networks 2012a, 2012b, 2012c, the latency, throughput, and cost of wide area network links 2013a, 2013b, and 2013c, 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.

Conclusion

[00187] 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.

[00188] 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.

[00189] 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.

[00190] 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.

[00191] 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.

[00192] 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.

[00193] 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 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.

[00194] The computer readable medium may also include non-transitory computer readable 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 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.

[00195] 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 device. However, other information transmissions may be between software modules and/or hardware modules in different physical devices.

[00196] Numerous modifications and variations of the present disclosure are possible in light of the above teachings.

TABLES

[00197] Table S1 below discloses crystallographic data collection and refinement statistics for the crystal structure of the 2bodx_43-b12 Fab complex.

2 bodx 43-b12 Fab	
PDB accession code	3RU8
Data collection	
Space Group	P2 ₁ 2 ₁ 2 ₁
Cell constants	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	90.5, 93.1, 94.4
α , β , γ (°)	90.0, 90.0, 90.0
Wavelength (Å)	1.00
Resolution (Å)	50.0-2.07 (2.11-2.07)
R _{merge}	10.3 (34.0)
Completeness (%)	99.4 (92.3)
Redundancy	3.1 (2.8)
Refinement	
Resolution (Å)	45.2-2.07 (2.12-2.07)
No. reflections	46486 (3201)
R _{work} / R _{free} (%)	16.9 / 20.2
No. atoms	
Protein	5512
Ligand/ion	24
Water	345
<i>B</i> -factors	
Protein	28.3
Ligand/ion	49.8
Water	32.0
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.156
Chiral volume (Å ³)	0.072
Estimated coordinate error (maximum likelihood e.s.u.) (Å)	0.096
Ramachandran	
Residues in favored regions (%)	90.8
Residues in allowed regions (%)	8.9
Residues in disallowed region (%)	0.2

Table S1

(values in parentheses are for highest-resolution shell)

[00198] Table S2 below discloses data for atomic contacts at the interface of the b12-gp120 complex (PDBID: 2NY7) and the b12-2bodx_43 complex. Residue numbering includes the letter for the insertion code and the antibody chain is presented in bold. The analysis was performed using the Contact application implemented under the CCP4 software suite.

2NY7		2bodb_43	
gp120	b12	2bodb_43	b12
Asn(280)	Arg(28H)		
Ala(281)	Arg(28H)		
		Tyr(110)	Tyr(98H), Trp(100H)
Ser(365)	Asn(31H)		
Gly(366)	Asn(31H), Phe(32H), Tyr(98H)	Gly(112)	Asn(31H), Phe(32H), Pro(97H), Tyr(98H)
Gly(367)	Asn(31H), Tyr(98H), Asn(100gH), Tyr(100hH)	Gly(113)	Asn(31H), Tyr(98H), Asn(100gH), Tyr(100hH)
Asp(368)	Asn(52H), Tyr(100hH)	Asp(114)	Asn(52H), Tyr(98H), Tyr(100hH), Val(33H)
Ile(371)	Asn(31H), Tyr(53H)	Ile(117)	Asn(31H)
		Gln(119)	Trp(100H)
		Trp(121)	Arg(28H)
		Leu(122)	Trp(100H)
Tyr(384)	Trp(100H)	His(151)	Ser(99H)
Asn(386)	Trp(100H)	Gln(155)	Trp(100H)
Arg(419)	Tyr(98H), Trp(100H)	Ala(158)	Trp(100H)
Lys(432)	Asn(56H)		
Thr(455)	Arg(28H)		
Arg(456)	Arg(28H)		
Asp(457)	Arg(28H)		
		Asp(72)	Arg(28H)
		Pro(75)	Asn(31H)
Arg(469)	Arg(28H)		
Gly(472)	Tyr(53H)		
Gly(473)	Ser(30H)	Gly(77)	Ser(30H), Tyr(54H)
Asp(474)	Tyr(53H)	Asp(78)	Pro(53H), Tyr(54H), Thr(73H)
Met(475)	Tyr(53H)	Met(79)	Tyr(54H), Asn(55H)
		Asn(217)	Arg(54L), Ala(55L), Gly(57L), Ile(58L)
		Glu(218)	Gly(57L)
		Asp(253)	Gln(1H)

Table S2

[00199] Table S3 below discloses data for hydrogen bonds at the interfaces of b12-gp120 (PDBid: 2NY7) and b12-2bodb_43 complexes. Residue numbering includes the letter for the insertion code and the antibody chain is presented in bold. The analysis was performed using the Contact application implemented under the CCP4 software suite.

2NY7			2bodb_43		
gp120	b12	Distance (Å)	2bodb_43	b12	Distance (Å)
Ala(281)-O	Arg(28H)-NH1	3.2			
			Tyr(110)-OH	Trp(100H)-N	3.10
Ser(365)-O	Asn(31H)-ND2	2.96			

Gly(366)-O	Tyr(98H)-N	2.78	Gly(112)-O	Tyr(98H)-N	2.89
Gly(367)-N	Asn(31H)-O	2.87	Gly(113)-N	Asn(31H)-O	2.88
			Gly(113)-O	Asn(100gH)-ND2	2.74
Asp(368)-OD1	Tyr(100hH)-OH	2.73	Asp(114)-OD1	Tyr(100hH)-OH	2.72
			Asp(114)-OD2	Asn(52H)-N2	3.18
			Asp(114)-OD2	Tyr(100hH)-OH	2.76
Arg(419)-NH1	Tyr(98H)-OH	3.56			
Arg(419)-NH1	Trp(100hH)-O	2.64			
Arg(419)-NH2	Trp(100hH)-O	3.49			
Lys(432)-NZ	Asn(56H)-ND2	3.11			
Thr(455)-OG1	Arg(28H)-NH1	3.39			
Arg(456)-O	Arg(28H)-NH2	3.39			
Asp(457)-OD1	Arg(28H)-NH2	3.52			
			Asp(72)-OD1	Arg(28H)-NH2	3.57
			Asp(72)-OD2	Arg(28H)-NH2	2.80
			Pro(75)-O	Asn(31H)-ND2	3.14
			Thr(76)-O	Ser(30H)-OG	2.76
			Gly(77)-O	Ser(30H)-OG	3.18
Asp(474)-OD1	Tyr(53H)-O	3.19	Asp(78)-OD2	Thr(73H)-OG1	3.48
Met(475)-N	Tyr(53H)-O	3.05	Met(79)-N	Tyr(54H)-O	2.99
			Trp(121)-NE1	Asn(31H)-OD1	2.76
			His(151)-NE2	Ser(103H)-OG	3.46
			Asn(217)-O	Gly(57L)-N	3.56
			Asn(217)-ND2	Ile(58L)-O	3.24
			Asn(217)-OD1	Ala(55L)-O	3.11

Table S3

CLAIMS

What is claimed:

1. A method for designing a functional protein scaffold comprising:
performing a scaffold search using a computing device, wherein the scaffold search includes a search of a protein structure library for a native structure similar to a structure of a functional motif, and wherein the native structure is supported by a protein scaffold;
removing the native structure from the protein scaffold to create a protein sequence gap; and
designing a functional protein scaffold by inserting the functional motif into the sequence gap of the protein scaffold.
2. The method of claim 1, further comprising changing, inserting, and/or removing an amino acid connecting segment flanking or within the functional motif.
3. The method of any of claims 1 and 2, further comprising: creating a computation-guided library accessible by the computing device, the computation-guided library comprising a plurality of library entries, and wherein at least one library entry of the plurality of library entries relates to the functional protein scaffold.
4. The method of claim 3, further comprising performing *in vitro* screening by screening the plurality of library entries for desired functional activity.
5. The method of any of claims 1-4, wherein performing a scaffold search comprises one or more of: Primary Loop matching, Secondary Loop matching, and a binding partner clash check.
6. The method of any of claims 1-5, wherein removing the native structure from the protein scaffold comprises one or more of: match reconstitution, build connecting segments by fragment insertion, computation sequence design, filtering, and human-guided design.
7. The method of any of claims 1-6, wherein creating a computation-guided library design comprises one or more of the following: generation of structural ensembles, computation sequence design and filter, explicit design recombination, conformational re-sampling of the interacting segment, and human-guided design and filtering.
8. The method of any of claims 1-7, wherein the protein scaffold is unrelated to the functional motif.

9. The method of any of claims 1-8, wherein the functional motif comprises a plurality of amino acid segments and wherein each of the plurality of amino acid segments are inserted into one or more protein sequence gaps of the protein scaffold.

10. The method of claim 9, wherein the plurality of amino acid segments are inserted into a plurality of protein sequence gaps of the protein scaffold.

11. The method of any of claims 1-10, wherein the functional motif is one of an epitope, binding site, and enzyme active site.

12. The method of any of claims 1-11, wherein the functional motif is immunogenic.

13. The method of any of claims 1-12, wherein the functional protein scaffold is one of a vaccine, an enzyme and a therapeutic.

14. The method of any of claims 1-13, wherein the functional protein scaffold binds to at least one of a biomolecule, inorganic molecule, and a surface.

15. The method of any of claims 1-14, wherein the biomolecule comprises an antibody, a cell surface marker, nucleic acid, an enzyme, and an inhibitor.

16. A functional protein scaffold for targeting mAb b12, comprising:
a protein scaffold having a first protein sequence gap and a second protein sequence gap;

a primary gp120 segment positioned in the first protein sequence gap and having a primary N-terminus and a primary C-terminus;

a secondary gp120 segment positioned in the second protein sequence gap and having a secondary N-terminus and a secondary C-terminus; and

one or more connecting segments flanking the primary and secondary gp120 segments at each of the segment termini.

17. An isolated polypeptide, comprising an amino acid sequence according to SEQ ID NO:1

18. The isolated polypeptide of claim 17, comprising an amino acid sequence according to SEQ ID NO:2.

19. The isolated polypeptide of any one of claims 17-18, wherein the polypeptide comprises a sequence selected from the group consisting of SEQ ID NOS:4-11.

20. An isolated polypeptide, comprising an amino acid sequence according to SEQ ID NO:12 or SEQ ID NO:13.

21. A virus-like particle (VLP) comprising the polypeptide of any one of claims 17-20.

22. An isolated nucleic acid encoding the polypeptide of any one of claims 17-20.
23. A recombinant expression vector comprising the isolated nucleic acid of claim 22 operatively linked to a promoter.
24. A recombinant host cell comprising the recombinant expression vector of claim 23.
25. A pharmaceutical composition, comprising the polypeptide of any one of claims 17-20, or the VLP of claim 21, and a pharmaceutically acceptable carrier.
26. A method for treating a immune deficiency virus (HIV) infection, comprising administering to a subject infected with an HIV infection an amount effective to treat the infection of the polypeptide of any one of claims 17-20, the VLP of claim 21 or the pharmaceutical composition of claim 25.
27. A method for limiting development of an HIV infection, comprising administering to a subject at risk of HIV infection an amount effective to limit development of an HIV infection of the polypeptide of any one of claims 17-20, the VLP of claim 21 or the pharmaceutical composition of claim 25.
28. A method for generating an immune response in a subject, comprising administering to the subject an amount effective to generate an immune response of the polypeptide of any one of claims 17-20, the VLP of claim 21 or the pharmaceutical composition of claim 25.
29. A pharmaceutical composition, comprising
 - (a) the isolated nucleic acid of claim 22, the recombinant expression vector of claim 23, and/or the recombinant host cell of claim 24; and
 - (b) a pharmaceutically acceptable carrier.
30. A method for monitoring an HIV infection or an HIV-induced disease in a subject and/or monitoring response of the subject to immunization by an HIV vaccine, comprising contacting the polypeptide of any one of claims 17-20, the VLP of claim 21 or the pharmaceutical composition of claim 25 with a bodily fluid from the subject and detecting HIV-binding antibodies in the bodily fluid of the subject.
31. The method of claim 16, wherein the bodily fluid comprises serum or whole blood.
32. A method for detecting HIV binding antibodies, comprising
 - (a) contacting the polypeptide of any one of claims 17-20, the VLP of claim 21 or the pharmaceutical composition of claim 25 with a composition comprising a candidate HIV

binding antibody under conditions suitable for binding of HIV antibodies to the polypeptide, VLP, or composition; and

(b) detecting HIV antibody complexes with the polypeptide, VLP, or composition.

33. The method of claim 32, further comprising isolating the HIV antibodies.

34. A method for producing HIV antibodies, comprising

(a) administering to a subject an amount effective to generate an antibody

response of the polypeptide of any one of claims 17-20, the VLP of claim 21 or the pharmaceutical composition of claim 25; and

(b) isolating antibodies produced by the subject.

35. An article of manufacture, comprising a physical computer-readable storage medium storing instructions that, upon execution by a processor, cause the processor to perform functions comprising:

performing a scaffold search, wherein the scaffold search includes a search of a protein structure library for a native structure similar to a structure of a functional motif, and wherein the native structure is supported by a protein scaffold;

removing the native structure from a representation of the protein scaffold to create a protein sequence gap; and

designing a functional protein scaffold by inserting a representation of the functional motif into a representation of the protein sequence gap of the protein scaffold.

36. The article of manufacture of claim 35, wherein the functions further comprise: changing, inserting, and/or removing a representation of an amino acid connecting segment flanking or within the functional motif.

37. The article of manufacture of either claim 35 or claim 36, wherein the functions further comprise:

creating a computation-guided library accessible by the computing device, the computation-guided library comprising a plurality of library entries, and wherein at least one library entry of the plurality of library entries relates to the functional protein scaffold.

38. The article of manufacture of any one of claims 35-37, wherein performing a scaffold search comprises one or more of: performing a match of a Primary Loop, performing a match of a Secondary Loop, and performing a binding partner clash check.

39. The article of manufacture of any one of claims 35-38, wherein removing the native structure from the representation of the protein scaffold comprises one or more of:

reconstituting a match, building connecting segments by fragment insertion, designing a sequence, filtering, and receiving an input related to human-guided design.

40. The article of manufacture of any one of claims 35-39, wherein creating a computation-guided library design comprises one or more of the following: generating structural ensembles, designing computation sequences, filtering the computation sequences, recombining designs, conformational re-sampling of a representation of an interacting segment, and receiving an input related to human-guided design and filtering.

41. The article of manufacture of any one of claims 35-40, wherein the protein scaffold is unrelated to the functional motif.

42. The article of manufacture of any one of claims 35-41, wherein the representation of the functional motif comprises a representation of a plurality of amino acid segments and wherein each representation of an amino acid segment is configured to be inserted into a representation of one or more protein sequence gaps of the protein scaffold.

43. The article of manufacture of claim 42, wherein the representation of the plurality of amino acid segments are configured to be inserted into a representation of a plurality of protein sequence gaps of the protein scaffold.

44. The article of manufacture of any one of claims 35-43, wherein the functional motif is one of an epitope, binding site, and enzyme active site.

45. The article of manufacture of any one of claims 35-44, wherein the functional motif is immunogenic.

46. The article of manufacture of any one of claims 35-45, wherein the functional protein scaffold is one of a vaccine, an enzyme and a therapeutic.

47. The article of manufacture of any one of claims 35-46, wherein the functional protein scaffold binds to at least one of a biomolecule, inorganic molecule, and a surface.

48. The article of manufacture of any one of claims 35-47, wherein the biomolecule comprises an antibody, a cell surface marker, nucleic acid, an enzyme, and an inhibitor.

49. A computing device, comprising:
a processor; and
data storage, storing instructions that, upon execution by the processor, cause the computing device to perform functions comprising:
performing a scaffold search, wherein the scaffold search includes a search of a protein structure library for a native structure similar to a structure of a functional motif, and wherein the native structure is supported by a protein scaffold,

removing the native structure from a representation of the protein scaffold to create a protein sequence gap, and

designing a functional protein scaffold by inserting a representation of the functional motif into a representation of the protein sequence gap of the protein scaffold.

50. The computing device of claim 49, wherein the functions further comprise: changing, inserting, and/or removing a representation of an amino acid connecting segment flanking or within the functional motif.

51. The computing device either claim 49 or claim 50, wherein the functions further comprise:

creating a computation-guided library accessible by the computing device, the computation-guided library comprising a plurality of library entries, and wherein at least one library entry of the plurality of library entries relates to the functional protein scaffold.

52. The computing device of any one of claims 49-51, wherein performing a scaffold search comprises one or more of: performing a match of a Primary Loop, performing a match of a Secondary Loop, and performing a binding partner clash check.

53. The computing device of any one of claims 49-52, wherein removing the native structure from the representation of the protein scaffold comprises one or more of: reconstituting a match, building connecting segments by fragment insertion, designing a sequence, filtering, and receiving an input related to human-guided design.

54. The computing device of any one of claims 49-53, wherein creating a computation-guided library design comprises one or more of the following: generating structural ensembles, designing computation sequences, filtering the computation sequences, recombining designs, conformational re-sampling of a representation of an interacting segment, and receiving an input related to human-guided design and filtering.

55. The computing device of any one of claims 49-54, wherein the protein scaffold is unrelated to the functional motif.

56. The computing device of any one of claims 49-55, wherein the representation of the functional motif comprises a representation of a plurality of amino acid segments and wherein each representation of an amino acid segment is configured to be inserted into a representation of one or more protein sequence gaps of the protein scaffold.

57. The computing device of any one of claims 49-56, wherein the representation of the plurality of amino acid segments are configured to be inserted into a representation of a plurality of protein sequence gaps of the protein scaffold.

58. The computing device of any one of claims 49-57, wherein the data storage is further configured to store at least part of the protein structure library.

59. The computing device of any one of claims 49-58, further comprising:
generating an output based on the functional protein scaffold.

60. The computing device of any one of claims 49-59, wherein the output comprises a sequence listing of at least part of the functional protein scaffold.

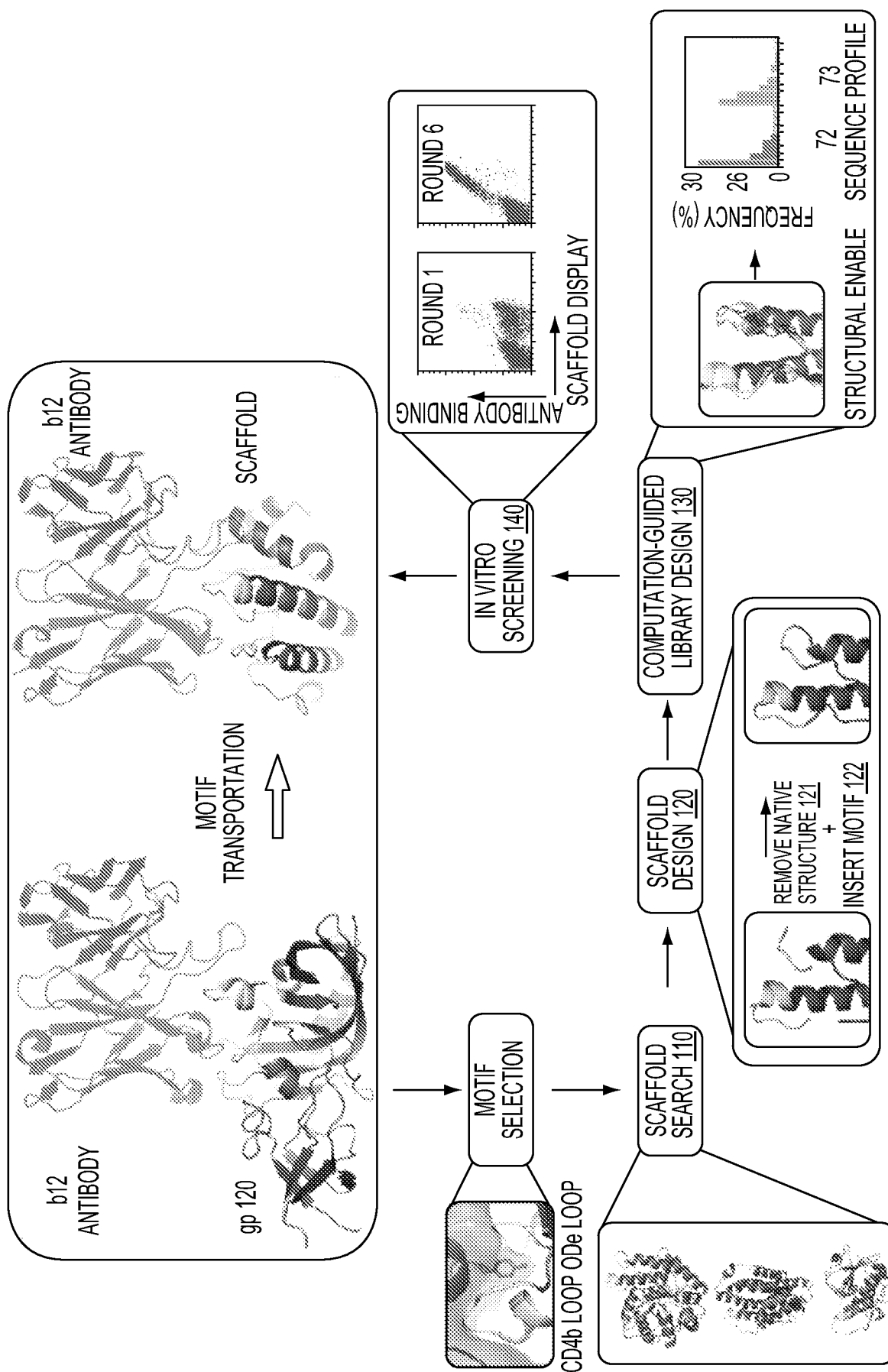


FIG. 1

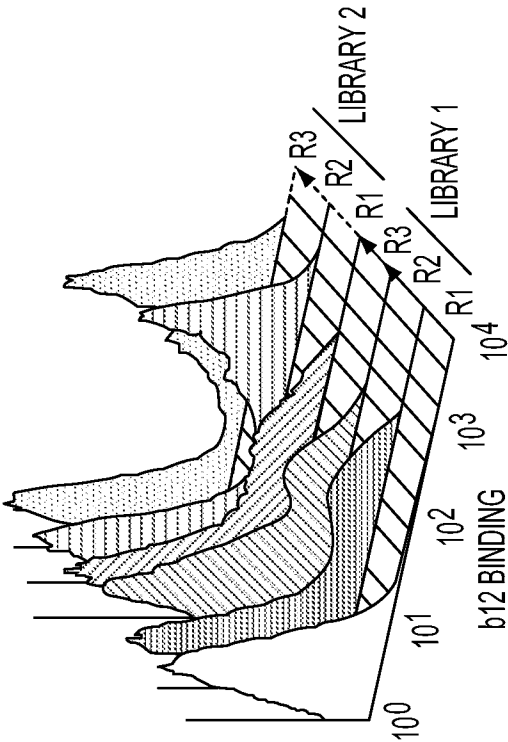


FIG. 2A

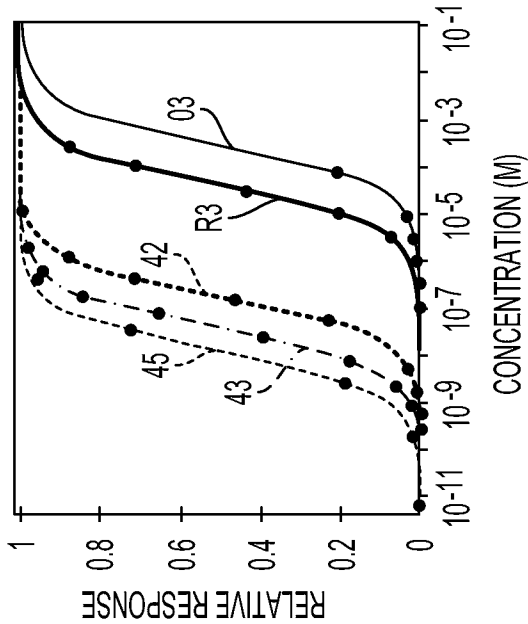


FIG. 2B

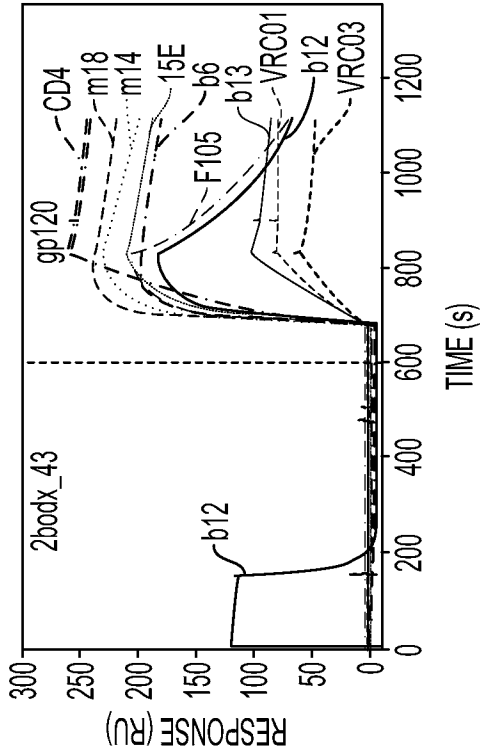


FIG. 2C

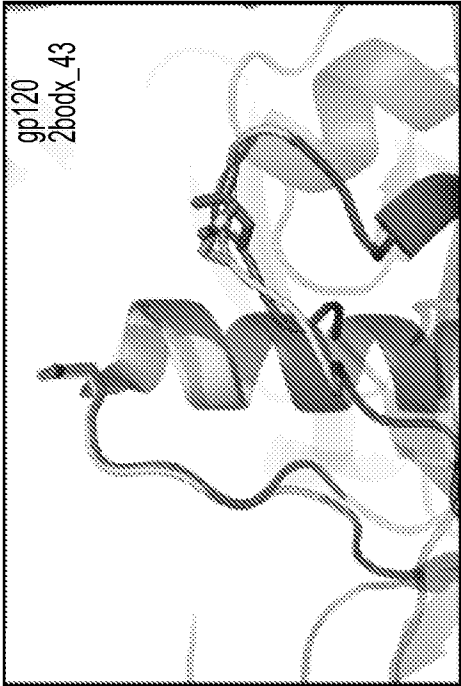


FIG. 3B

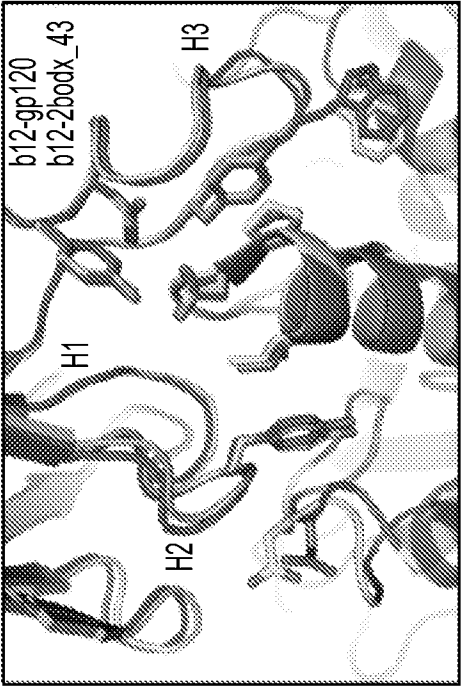


FIG. 3C

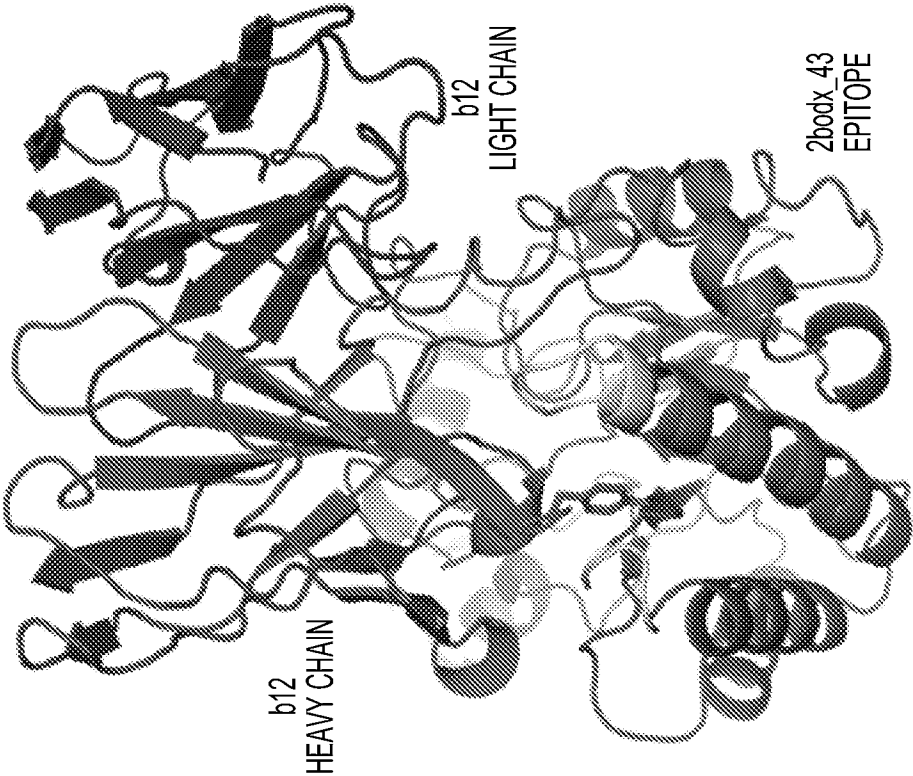


FIG. 3A

4/24

Multigraft Match

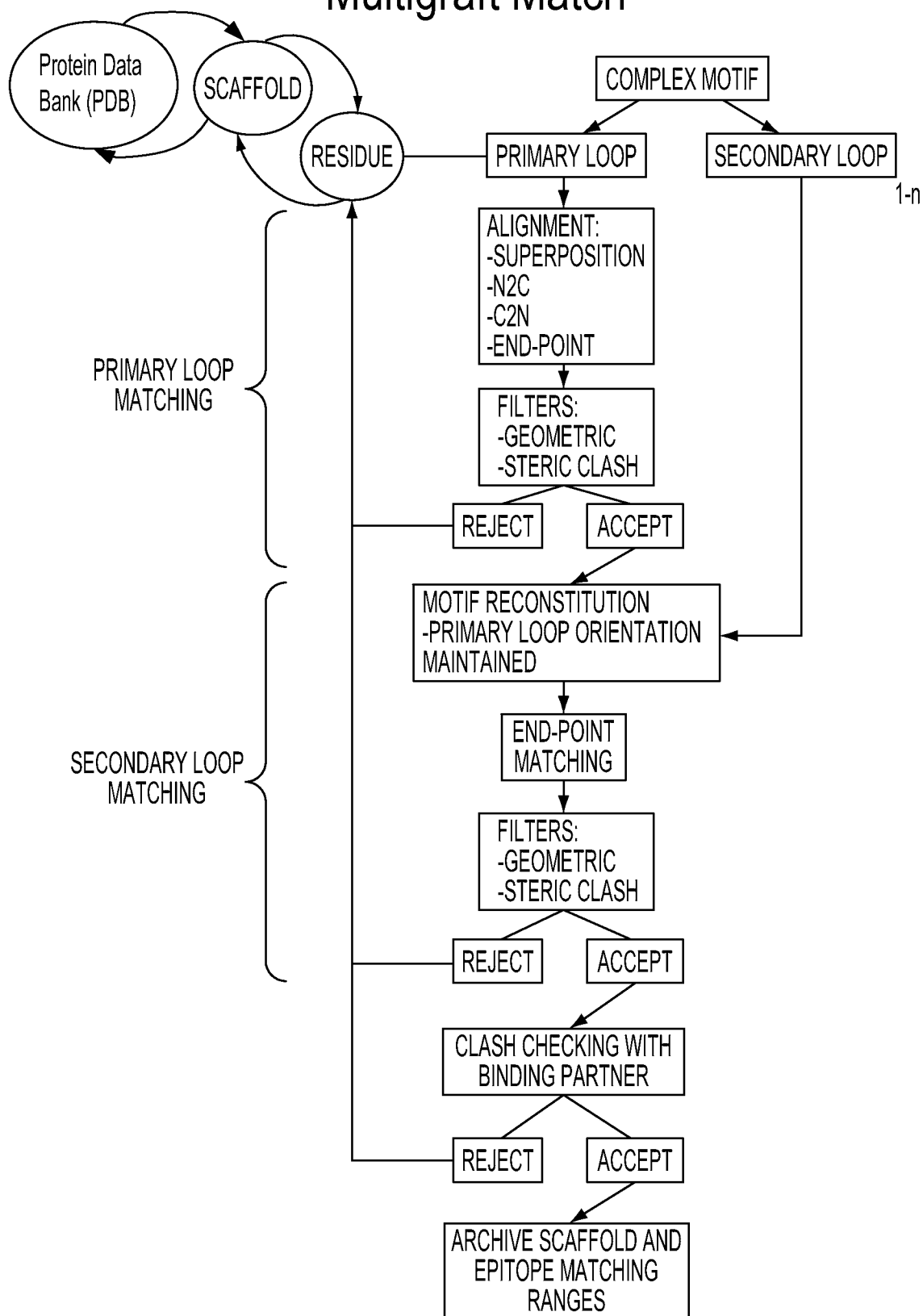


FIG. 4A

5/24

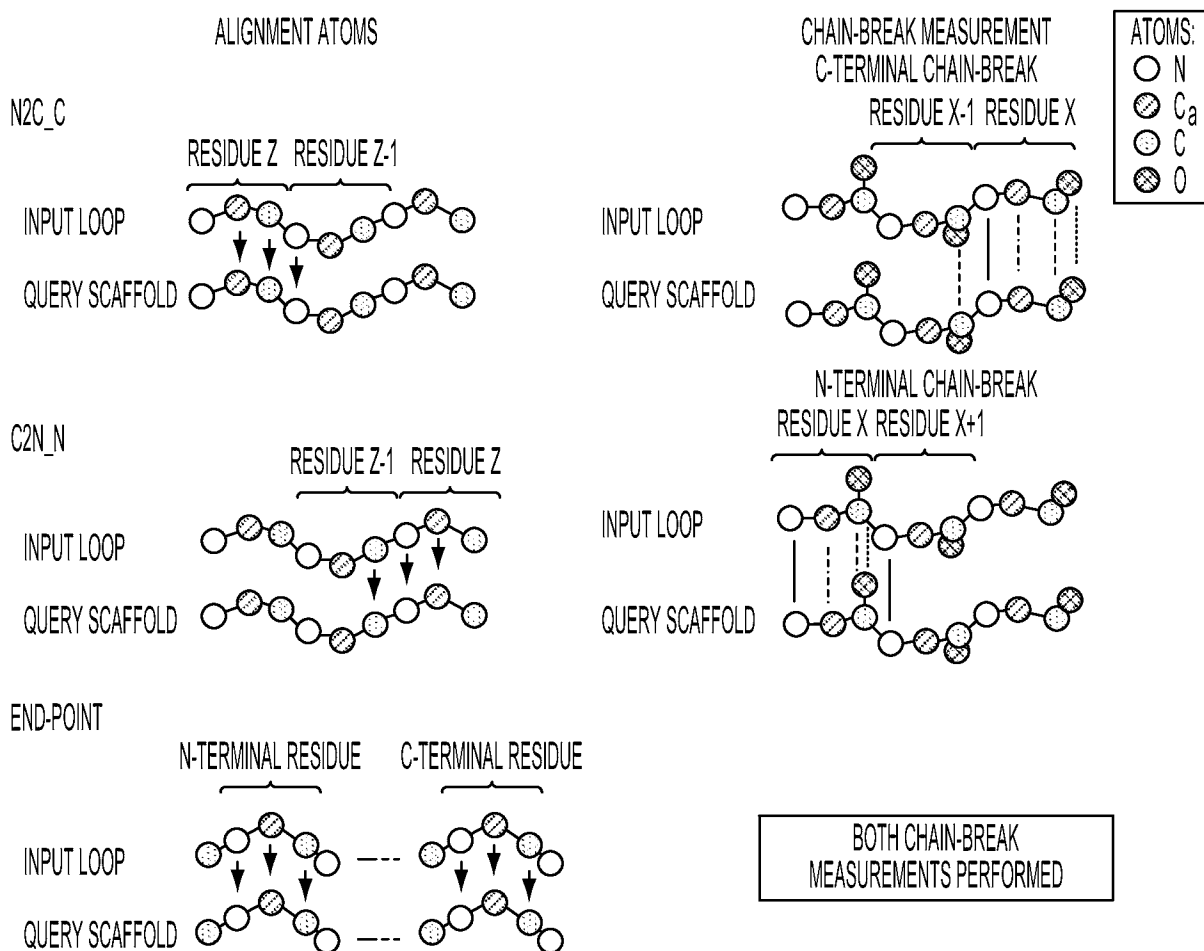
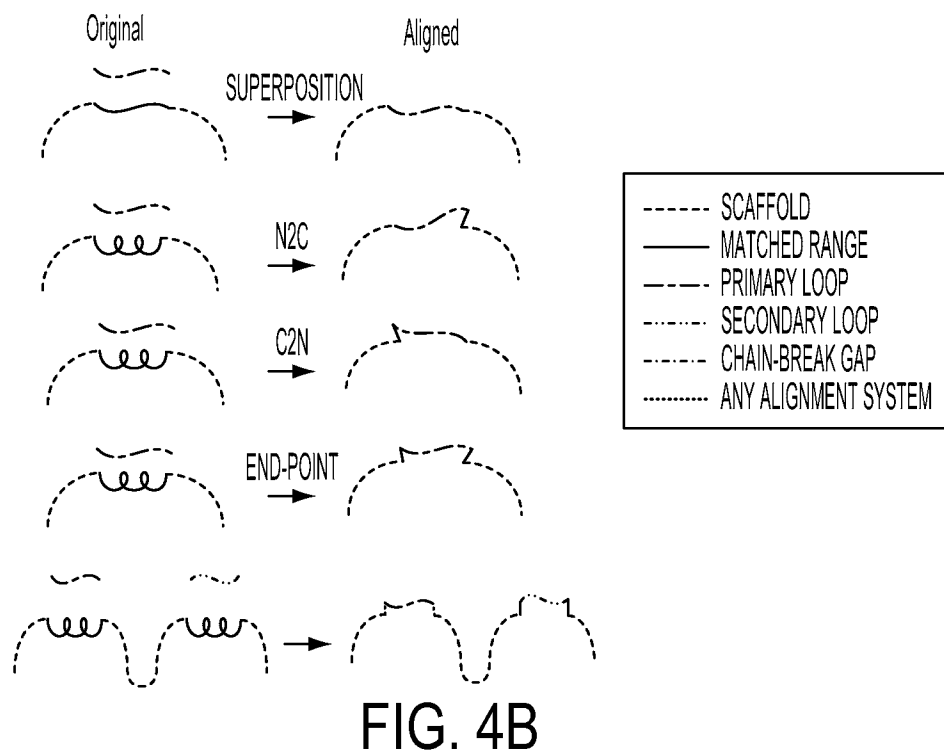


FIG. 4C

6/24

Multigraft Design

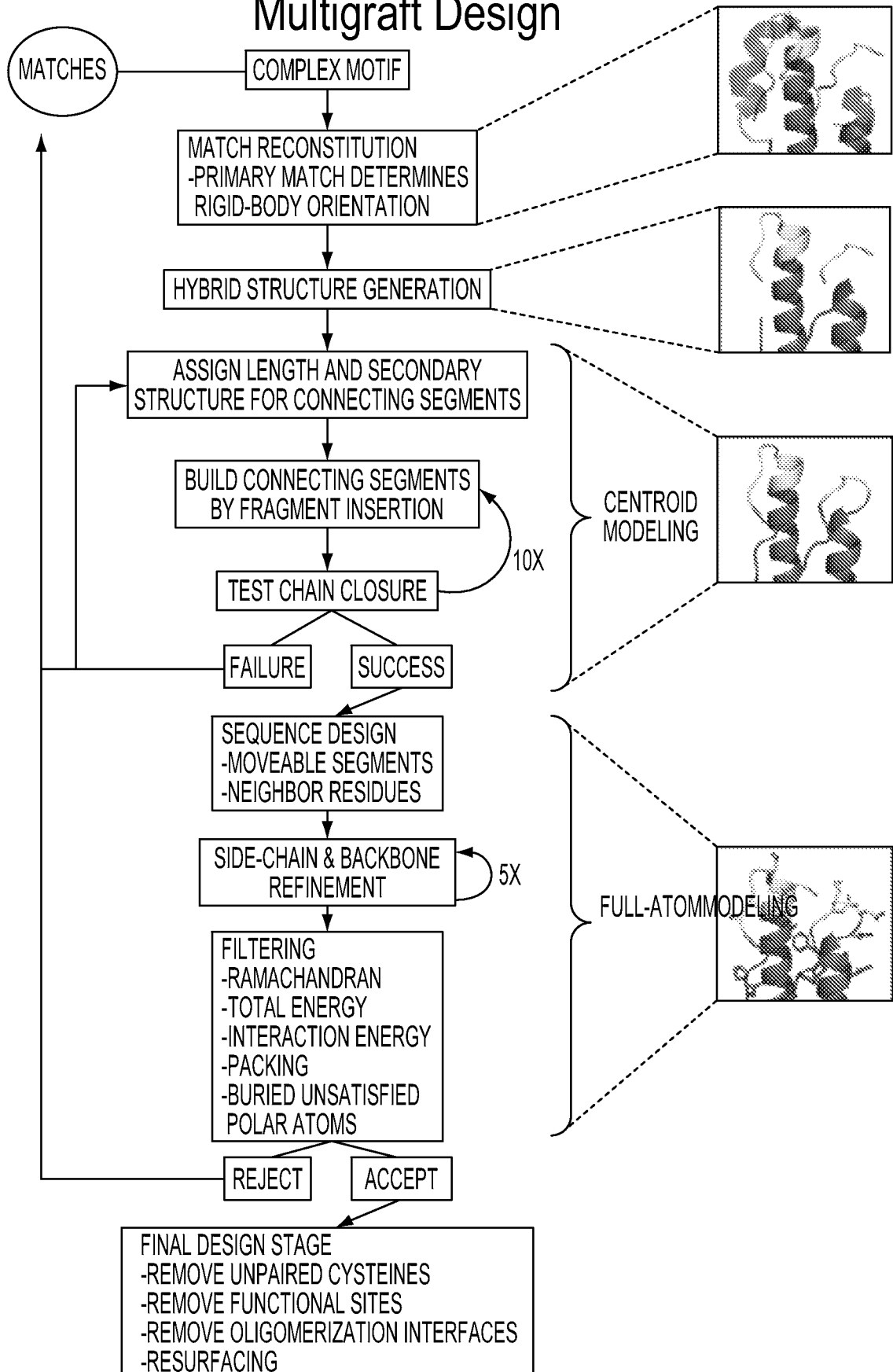


FIG. 5

7/24

NATIVE 2bodbx	DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTWFAHHNPGQITGQVDALMSAA	60
2bodbx_03	DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTW HNQ HNPQITGQVDALMSAA	60
2bodbx_43	DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTW HNQ HNPQITGQVDALMSAA	60
2bodbx_45	DSPFYVNPNMSSAEWVRNNPNDPRTPVIRDRIASVPQGTW HNQ HNPQITGQVDALMSAA	60
	*****. :*****	
	oooo ooo o o o oo o	
NATIVE 2bodbx	QAAGKIPILVVYNAPGRDCGNHSSGGAPSHSAYRSWIDEFAAGLKNRPAYIIVEPDLISL	120
2bodbx_03	QAAGKIPILVV NPD P---GDM DNG --- FE EKGQWIDEFAAGLKNRPAYIIV D P--- G	109
2bodbx_43	QAAGKIPILVVD VGP T---GDMS QG --- EA EKGQWIDEFAAGLKNRPAYIIV P --- L	109
2bodbx_45	QAAGKIPILVVD VGP T---GDMS QG --- EA EKGQWIDEFAAGLKNRPAYIIV P --- L	109
	***** * : ..* . :.***** *	
	o o o o o o oDe loop	
NATIVE 2bodbx	MSSCMQHVVQEVLETMAYAGKALKAGSSQARIYFDAGHSAWHSPAQMASWLQQADISNSA	180
2bodbx_03	G SGGDPEI AEAAWR FAAYAGKALKAGSSQARIYFDAGHSAWHSPAQMA AA LQRADISNSA	169
2bodbx_43	YSGGDPEIV QEW LRTVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMA AA LQRADISNSA	169
2bodbx_45	YSGGDPEIV QEW LRMVAYAGKALKAGSSQARIYFDAGHSAWHSPAQMA AA LQRADISNSA	169
	* . .: : . *****: **:*****	
	CD4b loop	
NATIVE 2bodbx	HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNEWCDPSGRAIGT	240
2bodbx_03	HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNE S CDPSGRAIGT	229
2bodbx_43	HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNE S CDPSGRAIGT	229
2bodbx_45	HGIATNTSNYRWTADDEVAYAKAVLSAIGNPSLRAVIDTSRNGNGPAGNE S CDPSGRAIGT	229
	***** *****	
NATIVE 2bodbx	PSTTNTGDPMIDAFWLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA	285 (SEQ. ID No: 15)
2bodbx_03	PSTTNTGDPMIDAFWLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA	274 (SEQ. ID No: 3)
2bodbx_43	PSTTNTGDPMIDAFWLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA	274 (SEQ. ID No: 4)
2bodbx_45	PSTTNTGDPMIDAFWLWIKLPGEADGCIAGAGQFVPQAAYEMAIAA	274 (SEQ. ID No: 5)

FIG. 6A

8/24

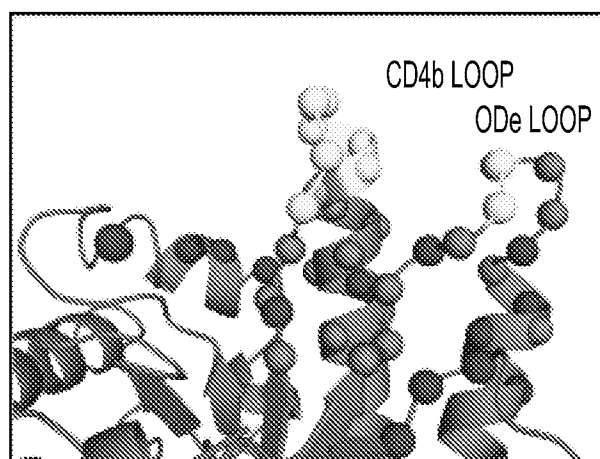


FIG. 6B

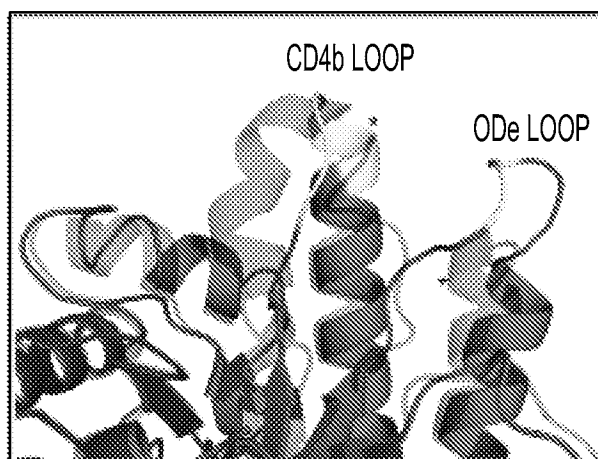


FIG. 6C

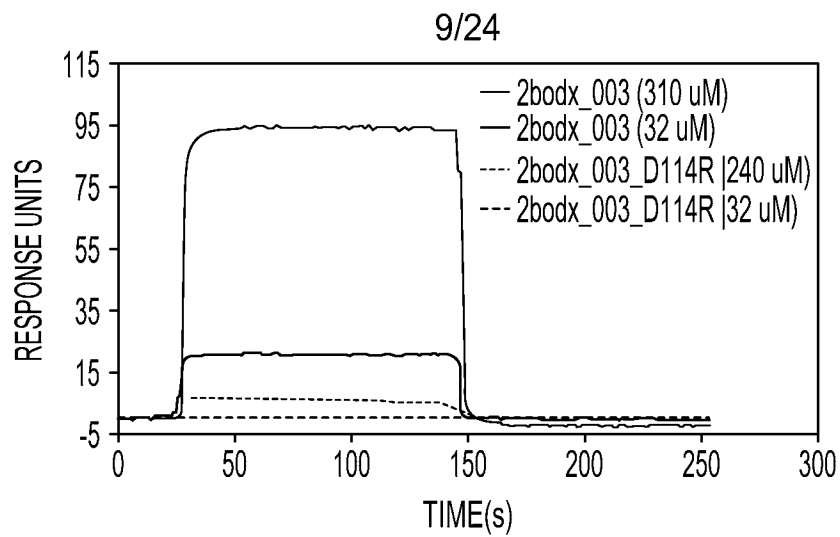


FIG. 7A

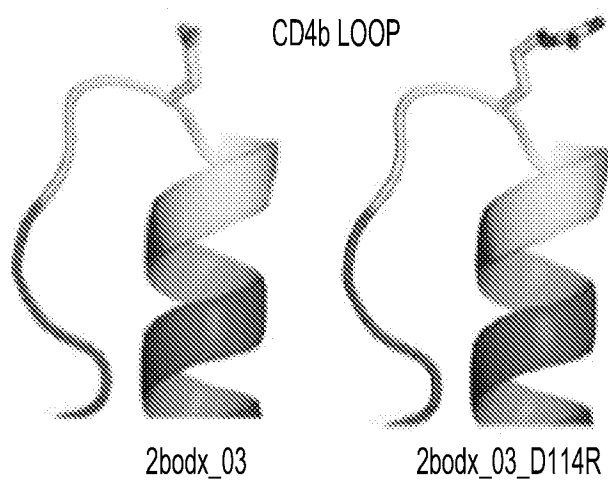


FIG. 7B

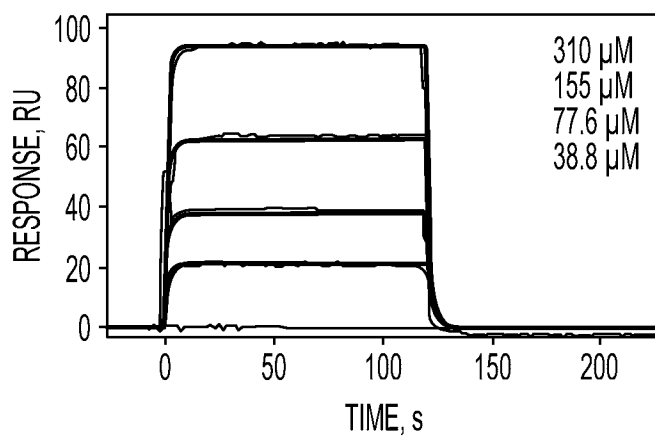


FIG. 7C

10/24

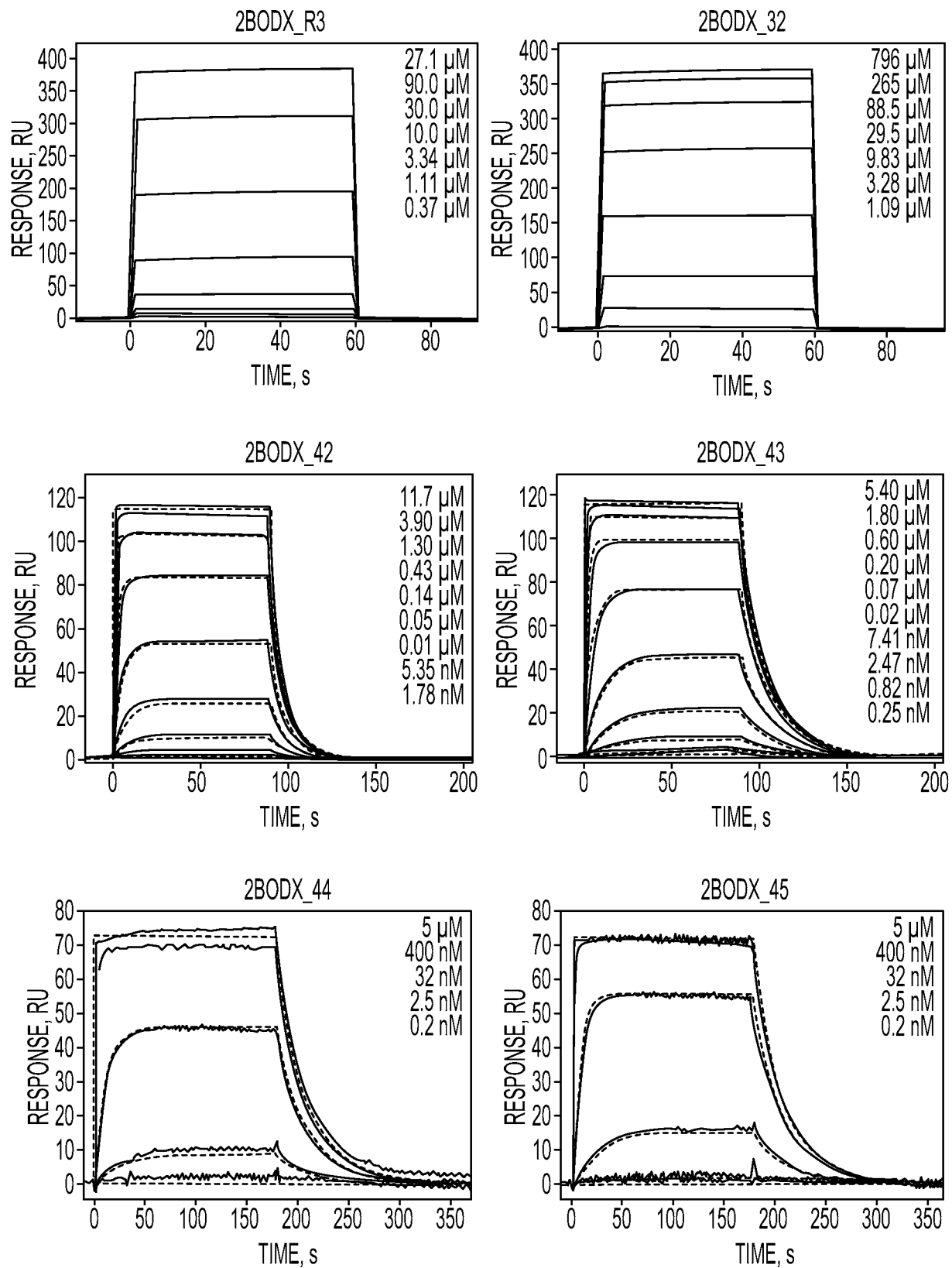


FIG. 8

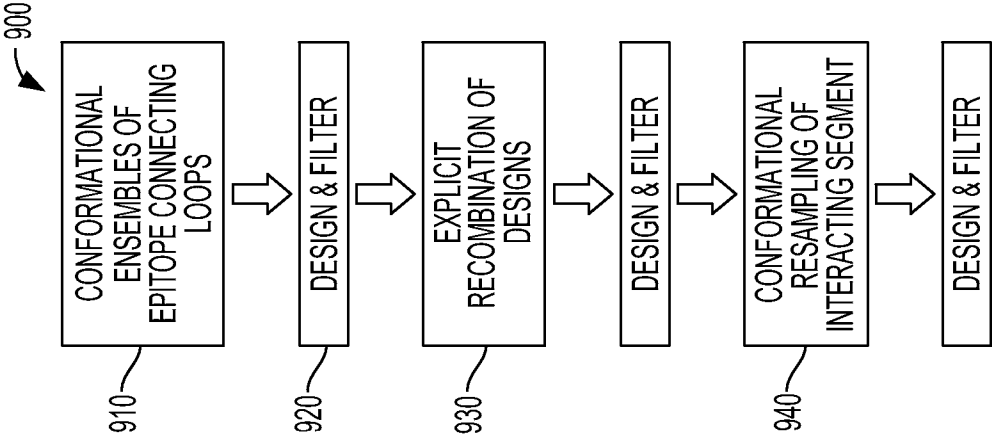


FIG. 9A

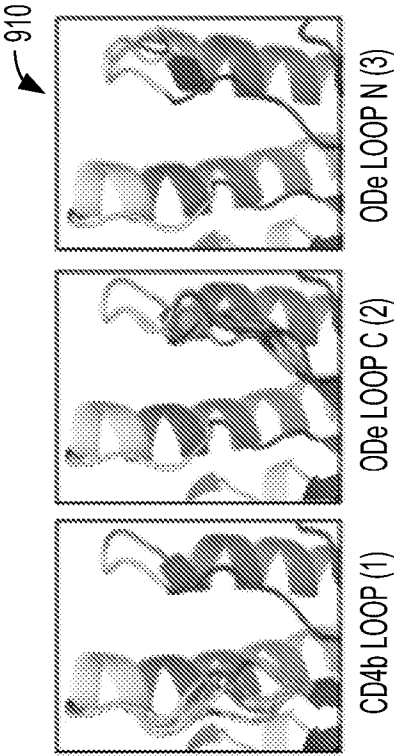


FIG. 9B

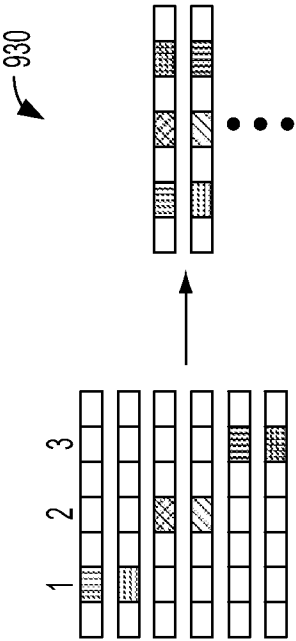
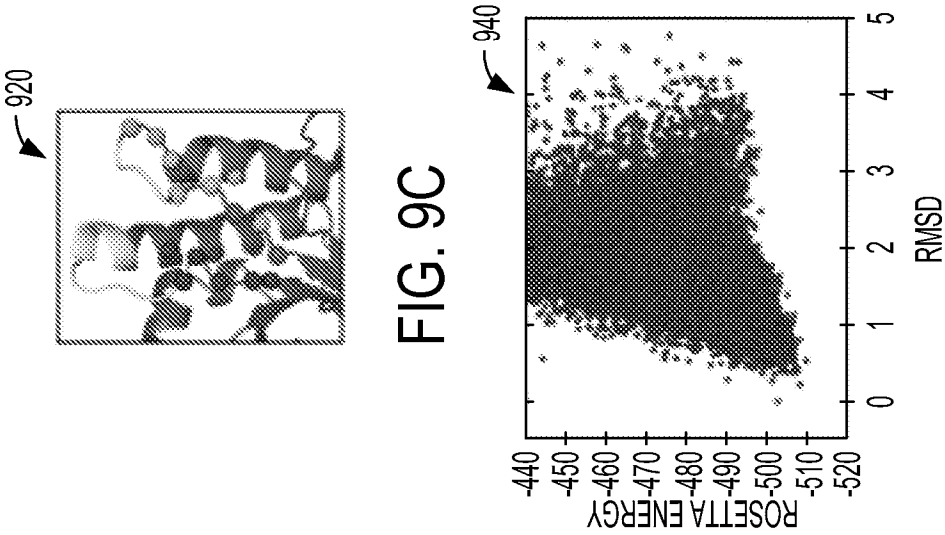


FIG. 9D



12/24

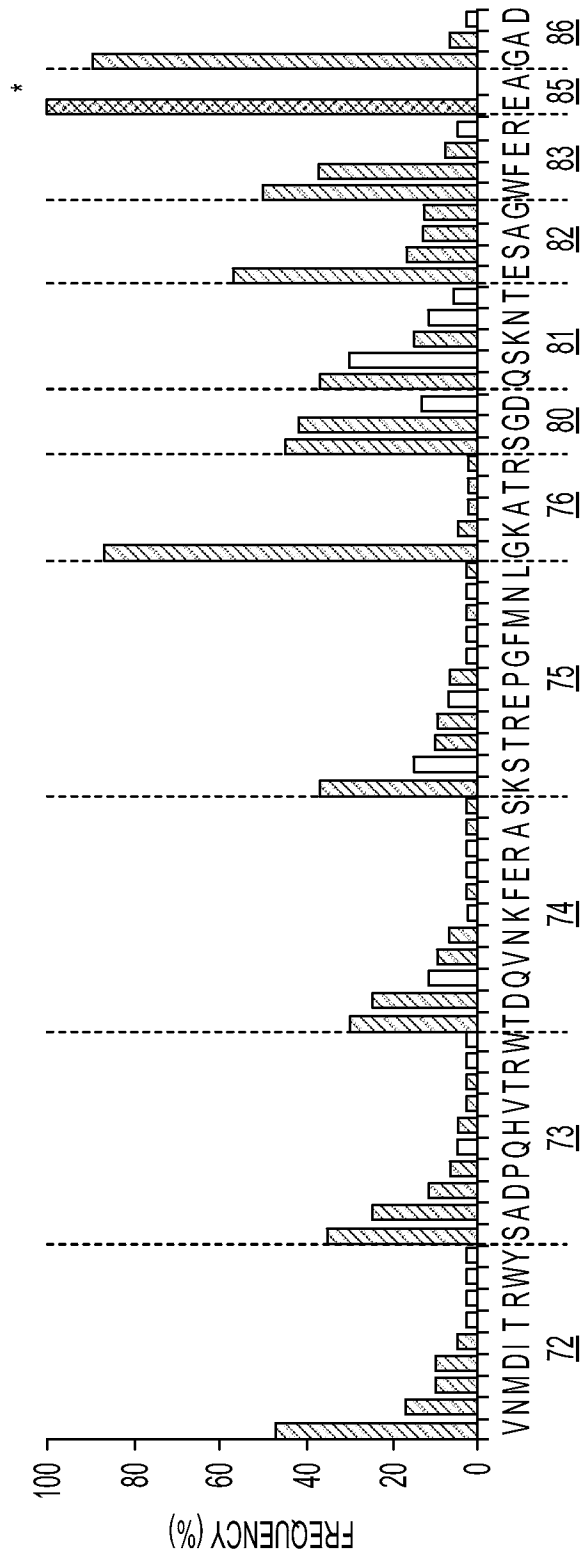


FIG. 10A

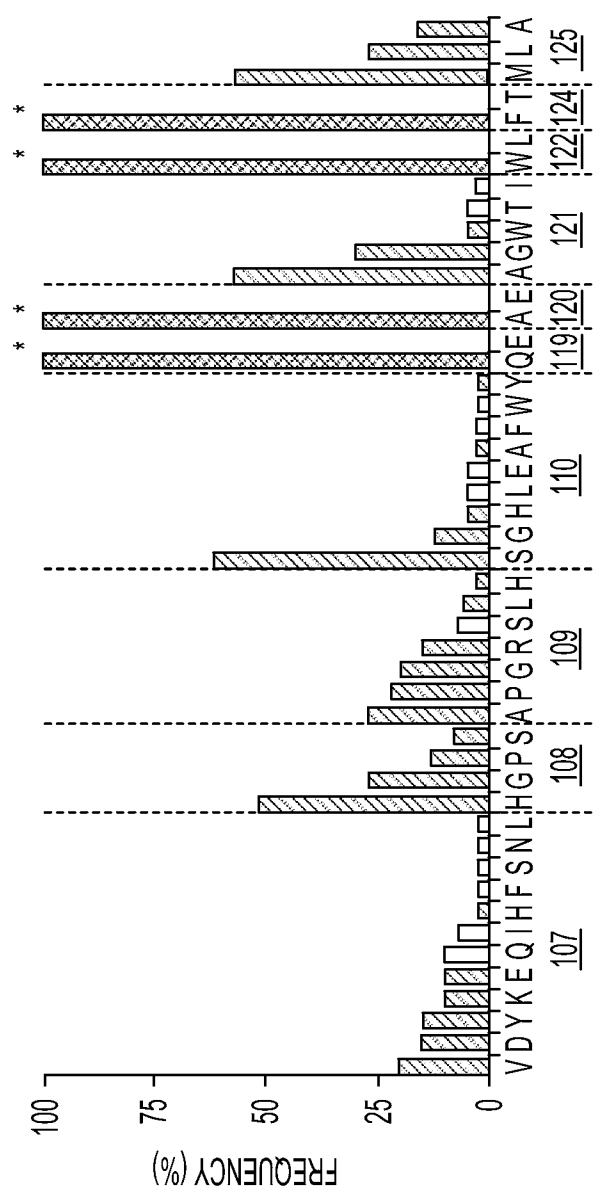


FIG. 10B

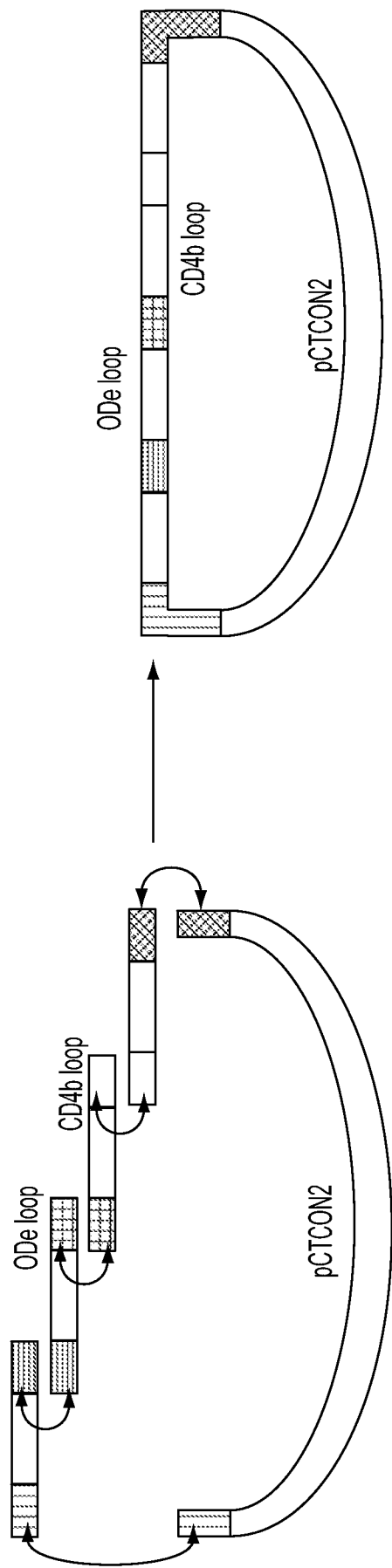


FIG. 11A

LIBRARY #	CD4b LOOP VARIANTS	ODe LOOP VARIANTS	THEORETICAL LIBRARY SIZE	EXPERIMENTAL LIBRARY SIZE
1	8	10^7	8×10^7	5×10^7
2	2.2×10^5	$>18^*$	$>4 \times 10^6$	7×10^6
3	3.8×10^2	4.3×10^3	1.6×10^6	5×10^6

FIG. 11B

Library 1

Position	72	73	74	75	76	80	81	82	83	85	86
2bod_03	N	P	D	P	G	D	N	G	F	E	G
ODe Loop diversity	D	V	G'	P	T	S	Q	G	E	A*	G
	V	S	T	K	G	G	K	S	F	E*	A
	N	A	D	T	K			E	W		
	M	D	V	R	A			A			
	I	P	N	L	R						
	E'	H	F	M							
	K'	T	A	Q'							
		L'	S								
			Y'								
			I'								
			C'								
# variants	4x10 ⁶										

Position	107	108	109	110	121	125
2bod_03	D	P	G	G	A	A
CD4b Loop diversity	D	P	G	G	A	A
	K	H	A	E	A	M
	Y	H	A	S	A	M
	Y	H	A	S	W	M
	Q	H	R	S	G	M
	E	S	L	H	G	M
	S	S	S	H	G	M
# variants	8					

Library 3

Position	75	76	83	86
2bod_03	P	G	F	G
OEe Loop diversity	G	A	A	A
	A	V	V	G
	P	L	L	V
	V	I	P	L
	L	F	E	F
	I	S	Q	Y
	R	T		I
	S	N		H
	T			N
# Variants	4.3x10 ³			

Position	109	120	124
2bod_03	G	A	F
CD4b Loop diversity	A	A	A
	I	V	I
	L	L	L
	M	P	M
	F	E	F
	S	Q	S
	T		T
	V		V
# Variants	3.8 x 10 ²		

FIG. 12A

16/24

Library 2

Position	# clones	72	73	74	75	76	80	81	82	83	85	86
2bod_03	-	N	P	D	P	G	D	N	G	F	E	G
ODe Loop diversity	11	N	L	Y	P	A	S	Q	E	E	A	A
	4	N	L	T	P	A	S	K	E	E	A	A
	3	E	A	D	T	R	G	Q	G	E	A	A
	3	D	A	S	P	T	S	K	E	W	E	K
	3	N	L	T	P	A	S	Q	E	E	A	A
	3	D	A	A	P	A	S	Q	G	E	A	A
	3	N	D	G	P	T	S	Q	A	W	A	A
	3	D	P	A	P	A	G	Q	E	E	E	A
	3	E	P	V	P	T	G	K	G	W	A	A
	2	E	P	D	P	T	S	K	G	F	A	G
	2	D	V	A	P	T	S	Q	E	E	A	A
	2	D	P	A	P	G	G	Q	A	E	E	A
	1	M	L	S	P	A	S	Q	E	F	A	A
	1	N	L	T	P	T	S	Q	E	F	A	A
	1	M	A	S	L	A	S	Q	G	E	E	G
	1	D	V	G	P	T	S	Q	G	E	A	G
	1	D	P	A	P	A	G	Q	E	E	A	A
	1	V	D	C	T	S	G	Q	A	E	A	A
		...										
# variants		> 18										

Position	107	108	109	110	119	120	121	122	124	125
2bod_03	D	P	G	G	E	A	A	W	F	A
CD4b Loop diversity	Y	P	L	Y	Q*	E*	W	L*	T*	V'
	V	H	A	S	E*	A*	A	W*	F*	M
	D	G	P	G			G			L
	K	S	G	H			S'			A
	E	R'	R	A						T'
	H	D'	H	P'						F'
	M'	Y'	D'	D'						
		C'	V'	R'						
		A'		C'						
# variants	2.2 x 10 ⁵									

FIG. 12B

Lib.	Round	Variant	Freq (%)	ODe Loop													CD4b Loop																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
				72	73	74	75	76	80	81	82	83	85	86	107	108	109	110	SGGDPEIA	E	A	A	W	F	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
		2bidx	-	N	P	D	P	G	GDM	D	N	G	F	E	G	D	P	G	G																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					

FIG. 13

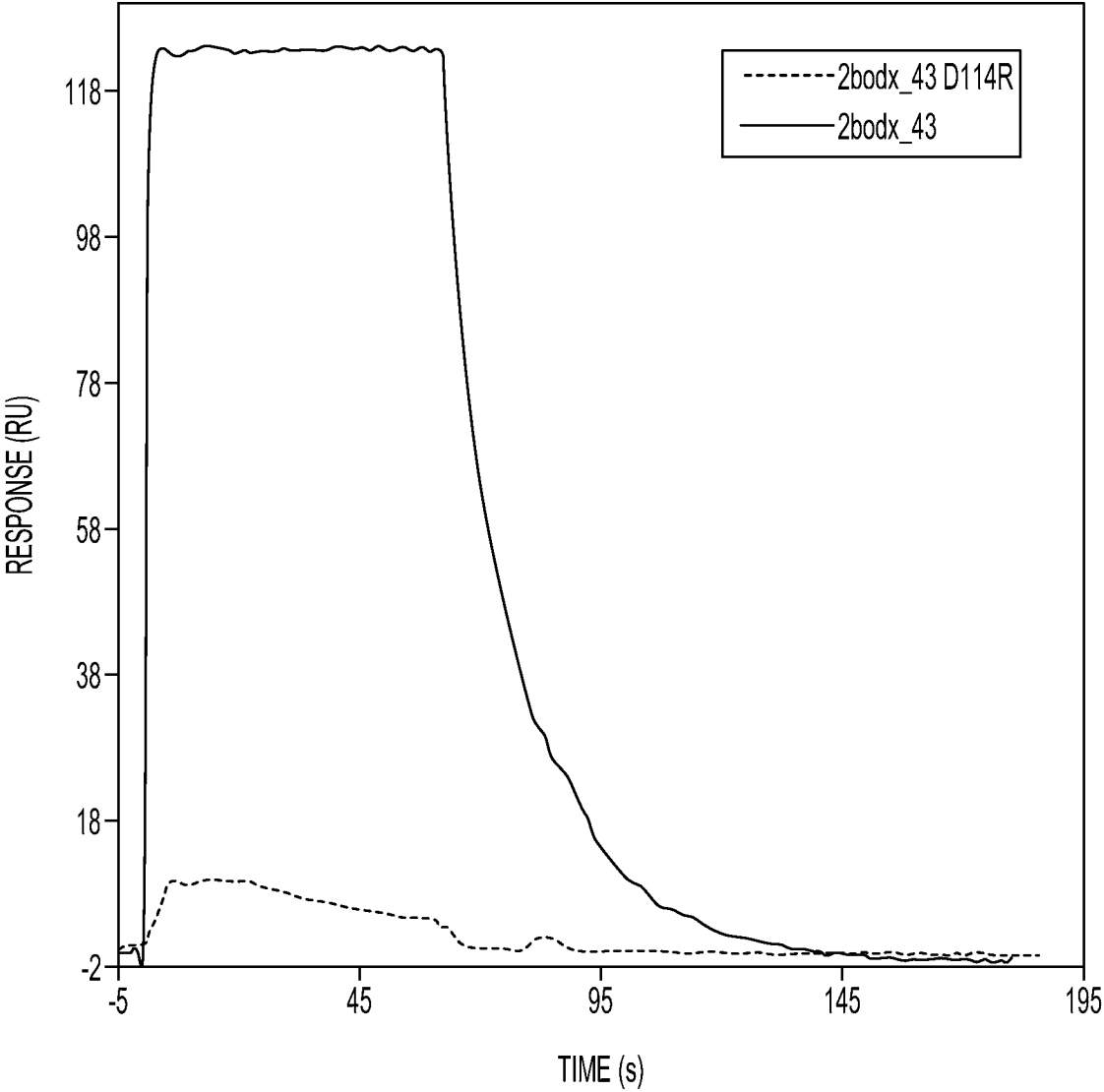


FIG. 14

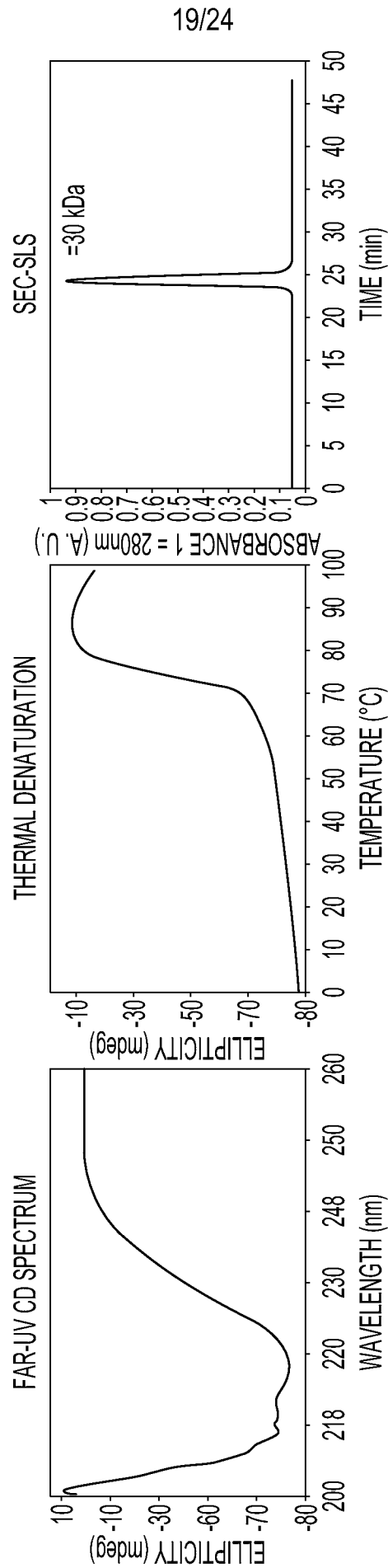


FIG. 15

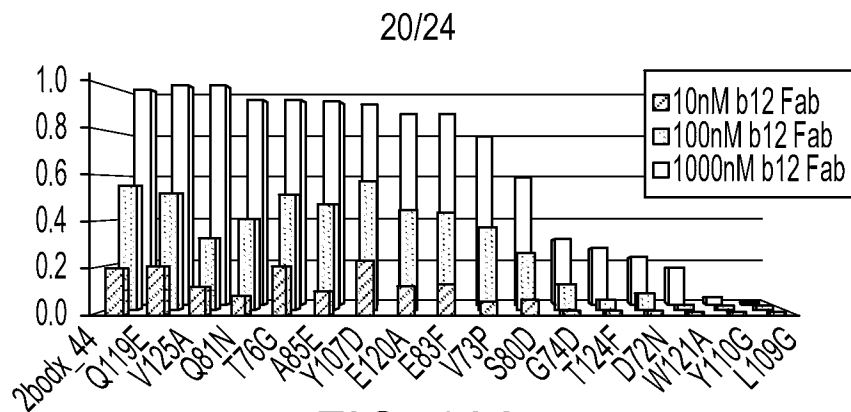


FIG. 16A

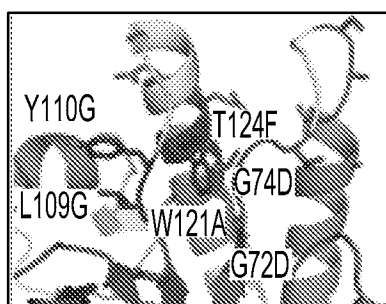


FIG. 16B

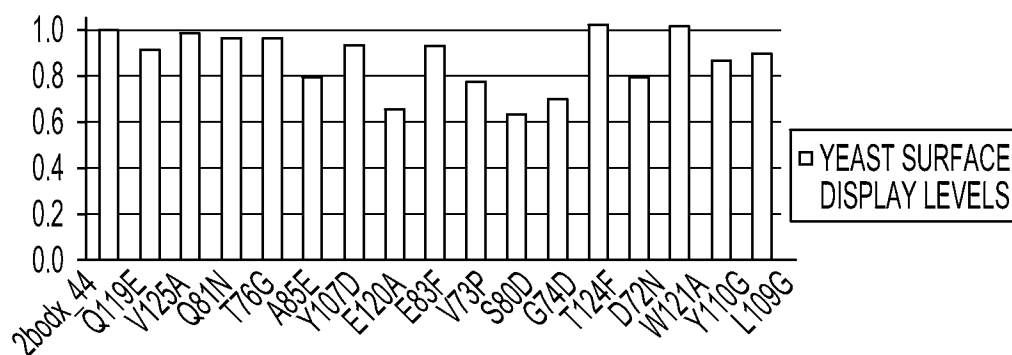


FIG. 16C

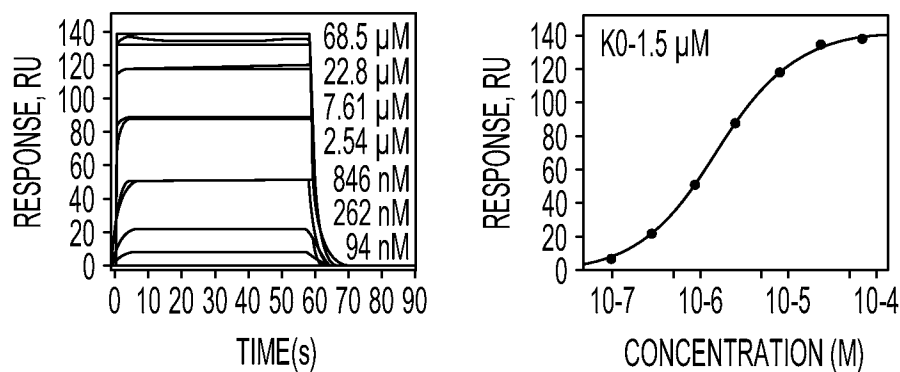


FIG. 16D

21/24

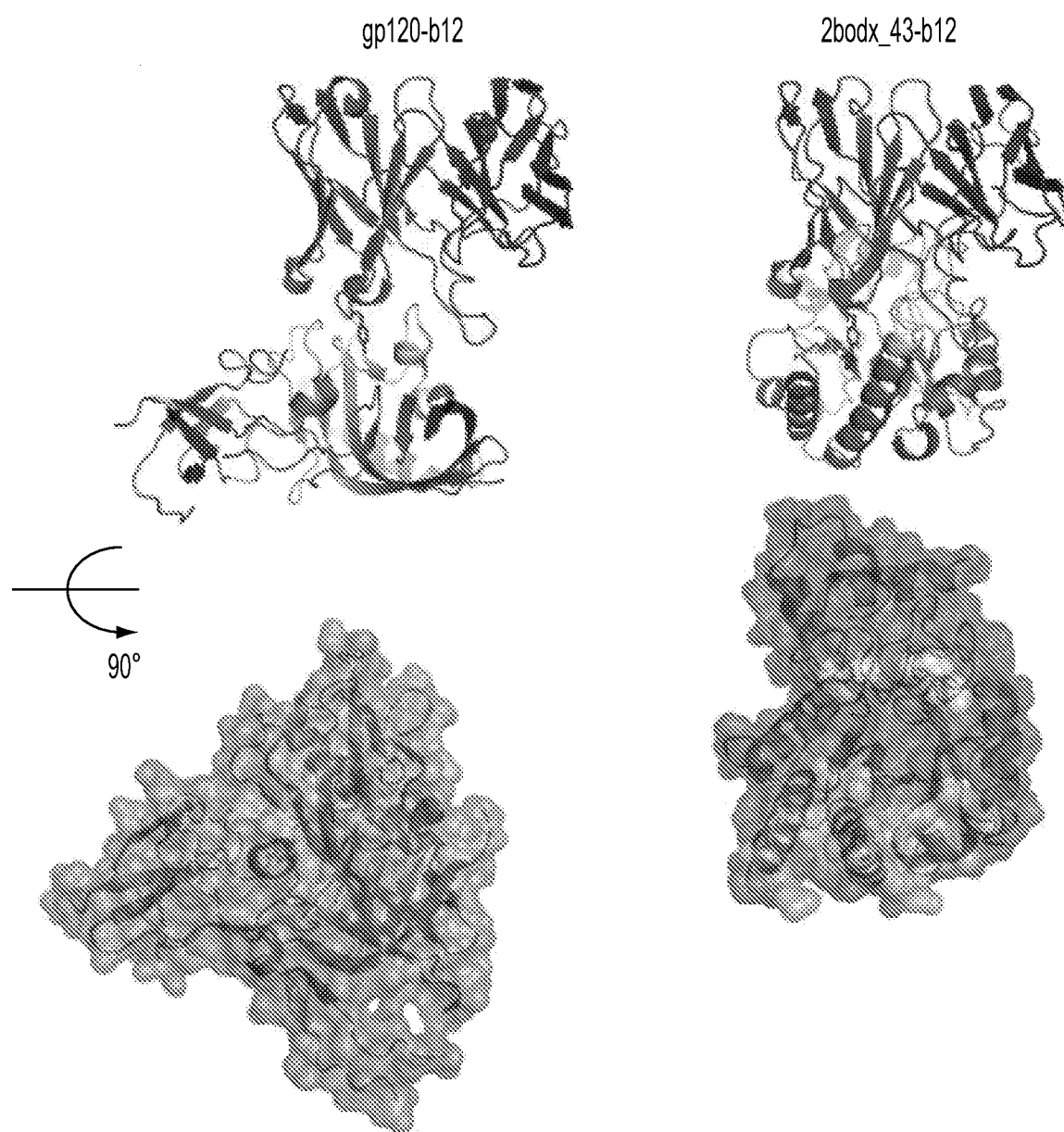


FIG. 17

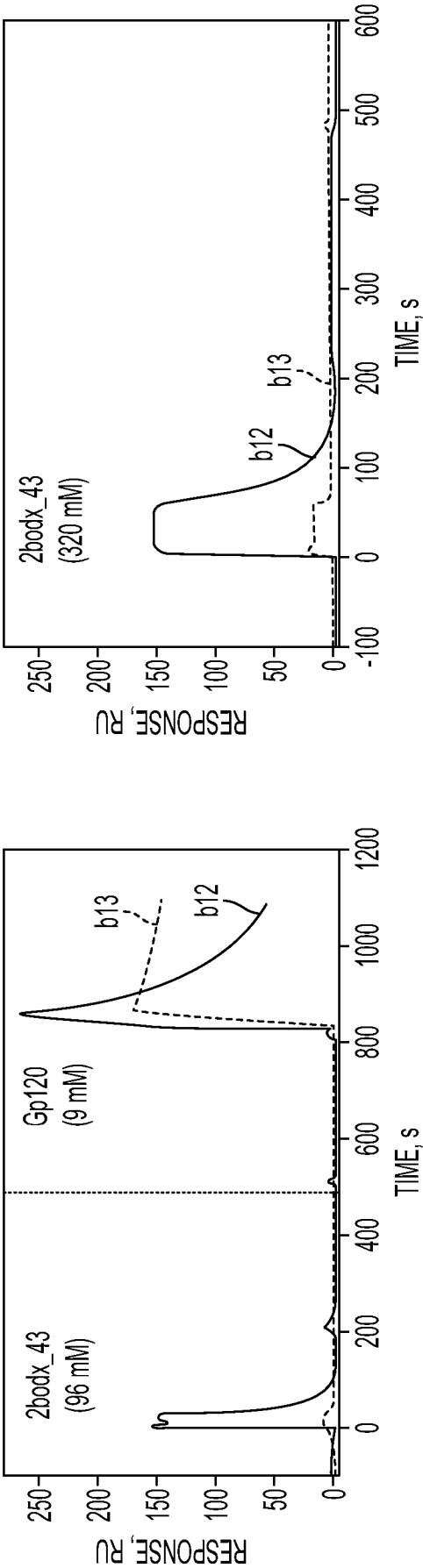


FIG. 18

23/24

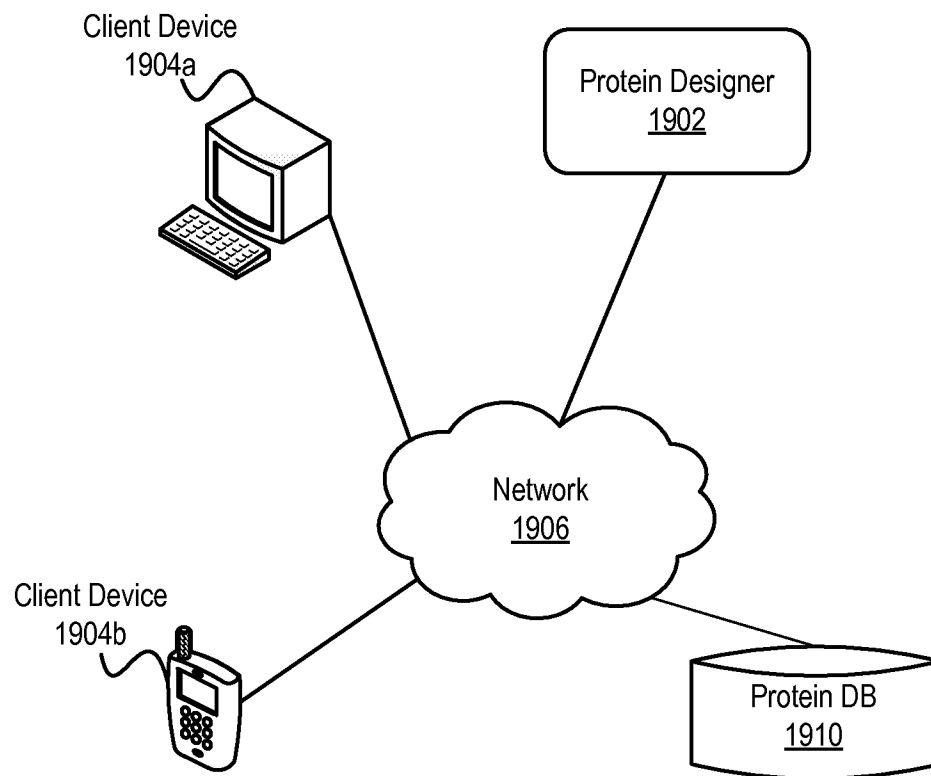


FIG. 19

24/24

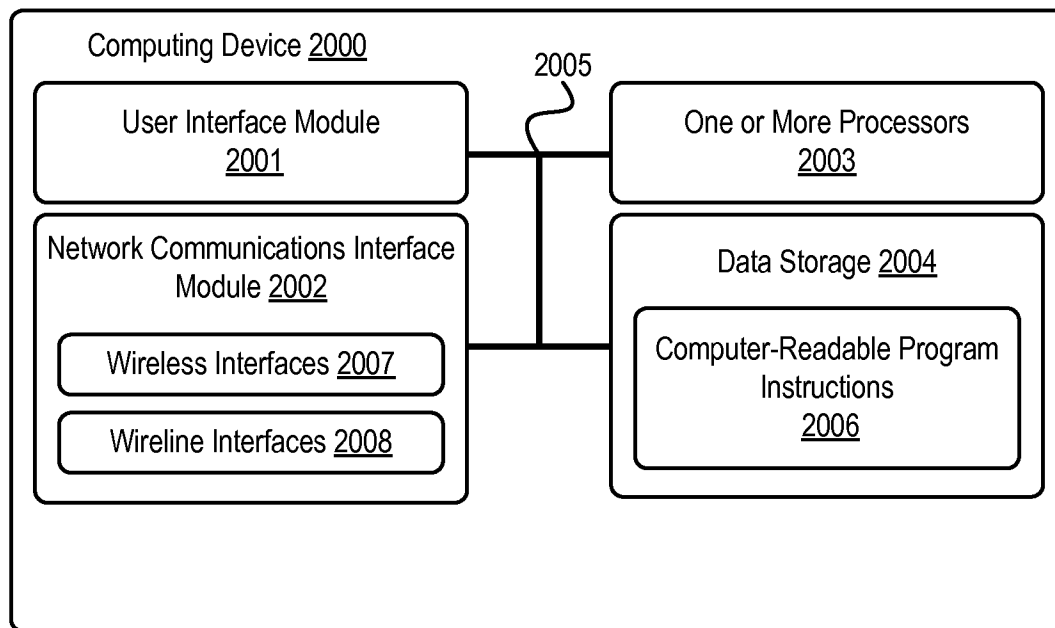


FIG. 20A

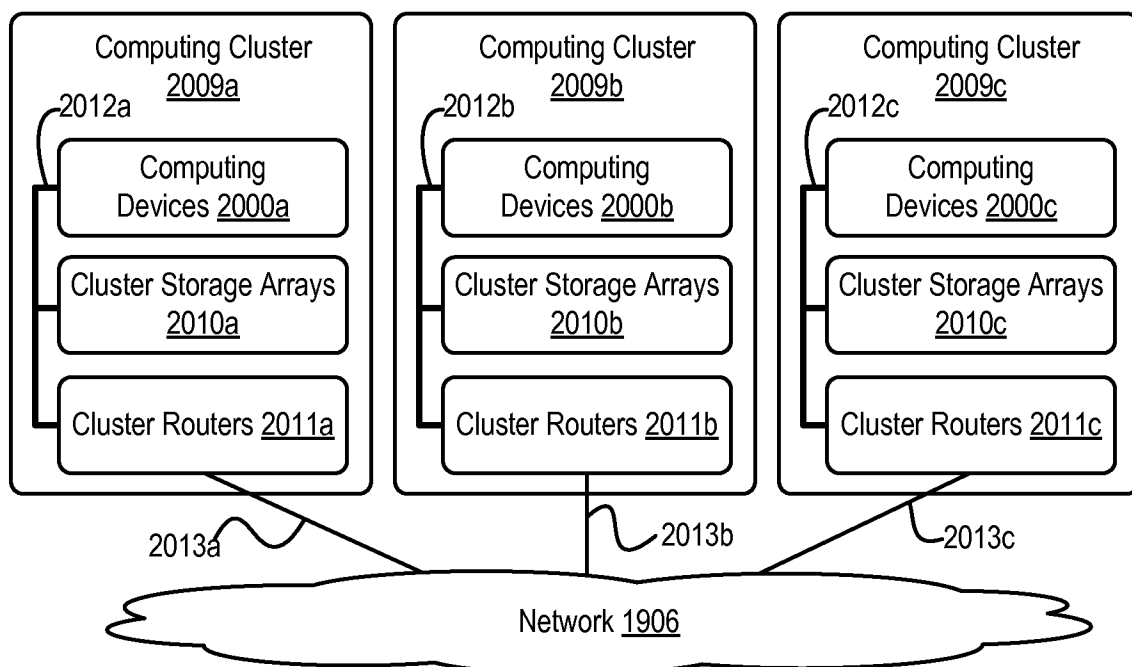


FIG. 20B