

GPCR BINDING PROTEINS AND SYNTHESIS THEREOF

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/556,863 filed on September 11, 2017 which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] G protein-coupled receptors (GPCRs) are implicated in a wide variety of diseases. Raising antibodies to GPCRs has been difficult due to problems in obtaining suitable antigen because GPCRs are often expressed at low levels in cells and are very unstable when purified. Thus, there is a need for improved agents for therapeutic intervention which target GPCRs.

INCORPORATION BY REFERENCE

[0003] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF SUMMARY

[0004] Provided herein are antibodies comprising a CDR-H3 comprising a sequence of any one of SEQ ID NOS: 2420 to 2436. Provided herein are antibodies comprising a CDR-H3 comprising a sequence of any one of SEQ ID NOS: 2420 to 2436; and wherein the antibody is a monoclonal antibody, a polyclonal antibody, a bi-specific antibody, a multispecific antibody, a grafted antibody, a human antibody, a humanized antibody, a synthetic antibody, a chimeric antibody, a camelized antibody, a single-chain Fvs (scFv), a single chain antibody, a Fab fragment, a F(ab')₂ fragment, a Fd fragment, a Fv fragment, a single-domain antibody, an isolated complementarity determining region (CDR), a diabody, a fragment comprised of only a single monomeric variable domain, disulfide-linked Fvs (sdFv), an intrabody, an anti-idiotypic (anti-Id) antibody, or ab antigen-binding fragments thereof. Provided herein are antibodies wherein the VH domain is IGHV1-18, IGHV1-69, IGHV1-8 IGHV3-21, IGHV3-23, IGHV3-30/33rn, IGHV3-28, IGHV3-74, IGHV4-39, or IGHV4-59/61. Provided herein are antibodies, wherein the VL domain is IGKV1-39, IGKV1-9, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV4-1, IGLV1-51, or IGLV2-14. Provided herein are methods of inhibiting GLP1R activity, comprising administering the antibodies as described herein. Provided herein are methods for treatment of a metabolic disorder, comprising administering to a subject in need thereof the antibodies as described herein. In some instances, the antibody comprises a CDR-H3 comprising a sequence of any one of SEQ ID NOS: 2420 to 2436. Provided herein are methods for

treatment of a metabolic disorder, wherein the metabolic disorder is Type II diabetes, or obesity. Provided herein are nucleic acids encoding for a protein comprising a sequence of any one of SEQ ID NOS: 2420 to 2436.

[0005] Provided herein are nucleic acid libraries comprising a plurality of nucleic acids, wherein each nucleic acid encodes for a sequence that when translated encodes for an immunoglobulin scaffold, wherein the immunoglobulin scaffold comprises a CDR-H3 loop that comprises a GPCR binding domain, and wherein each nucleic acid comprises a sequence encoding for a sequence variant of the GPCR binding domain. Provided herein are nucleic acid libraries, wherein a length of the CDR-H3 loop is about 20 to about 80 amino acids. Provided herein are nucleic acid libraries, wherein a length of the CDR-H3 loop is about 80 to about 230 base pairs. Provided herein are nucleic acid libraries, wherein the immunoglobulin scaffold further comprises one or more domains selected from variable domain, light chain (VL), variable domain, heavy chain (VH), constant domain, light chain (CL), and constant domain, heavy chain (CH). Provided herein are nucleic acid libraries, wherein the VH domain is IGHV1-18, IGHV1-69, IGHV1-8 IGHV3-21, IGHV3-23, IGHV3-30/33m, IGHV3-28, IGHV3-74, IGHV4-39, or IGHV4-59/61. Provided herein are nucleic acid libraries, wherein the VL domain is IGKV1-39, IGKV1-9, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV4-1, IGLV1-51, or IGLV2-14. Provided herein are nucleic acid libraries, wherein a length of the VH domain is about 90 to about 100 amino acids. Provided herein are nucleic acid libraries, wherein a length of the VL domain is about 90 to about 120 amino acids. Provided herein are nucleic acid libraries, wherein a length of the VH domain is about 280 to about 300 base pairs. Provided herein are nucleic acid libraries, wherein a length of the VL domain is about 300 to about 350 base pairs. Provided herein are nucleic acid libraries, wherein the library comprises at least 105 non-identical nucleic acids. Provided herein are nucleic acid libraries, wherein the immunoglobulin scaffold comprises a single immunoglobulin domain. Provided herein are nucleic acid libraries, wherein the immunoglobulin scaffold comprises a peptide of at most 100 amino acids. Provided herein are vector libraries comprising nucleic acid libraries as described herein. Provided herein are cell libraries comprising nucleic acid libraries as described herein.

[0006] Provided herein are nucleic acid libraries comprising a plurality of nucleic acids, wherein each nucleic acid encodes for a sequence that when translated encodes a GPCR binding domain, and wherein each nucleic acid comprises sequence encoding for a different GPCR binding domain about 20 to about 80 amino acids. Provided herein are nucleic acid libraries, wherein a length of the GPCR binding domain is about 80 to about 230 base pairs. Provided herein are nucleic acid libraries, wherein the GPCR binding domain is designed based on

conformational ligand interactions, peptide ligand interactions, small molecule ligand interactions, extracellular domains of GPCRs, or antibodies that target GPCRs. Provided herein are vector libraries comprising nucleic acid libraries as described herein. Provided herein are cell libraries comprising nucleic acid libraries as described herein.

[0007] Provided herein are protein libraries comprising a plurality of proteins, wherein each of the proteins of the plurality of proteins comprise an immunoglobulin scaffold, wherein the immunoglobulin scaffold comprises a CDR-H3 loop that comprises a sequence variant of a GPCR binding domain. Provided herein are protein libraries, wherein a length of the CDR-H3 loop is about 20 to about 80 amino acids. Provided herein are protein libraries, wherein the immunoglobulin scaffold further comprises one or more domains selected from variable domain, light chain (VL), variable domain, heavy chain (VH), constant domain, light chain (CL), and constant domain, heavy chain (CH). Provided herein are protein libraries, wherein the VH domain is IGHV1-18, IGHV1-69, IGHV1-8 IGHV3-21, IGHV3-23, IGHV3-30/33m, IGHV3-28, IGHV3-74, IGHV4-39, or IGHV4-59/61. Provided herein are protein libraries, wherein the VL domain is IGKV1-39, IGKV1-9, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV4-1, IGLV1-51, or IGLV2-14. Provided herein are protein libraries, wherein a length of the VH domain is about 90 to about 100 amino acids. Provided herein are protein libraries, wherein a length of the VL domain is about 90 to about 120 amino acids. Provided herein are protein libraries, wherein the plurality of proteins is used to generate a peptidomimetic library. Provided herein are protein libraries, wherein the protein library comprises peptides. Provided herein are protein libraries, wherein the protein library comprises immunoglobulins. Provided herein are protein libraries, wherein the protein library comprises antibodies. Provided herein are cell libraries comprising protein libraries as described herein.

[0008] Provided herein are protein libraries comprising a plurality of proteins, wherein the plurality of proteins comprises sequence encoding for different GPCR binding domains, and wherein the length of each GPCR binding domain is about 20 to about 80 amino acids. Provided herein are protein libraries, wherein the protein library comprises peptides. Provided herein are protein libraries, wherein the protein library comprises immunoglobulins. Provided herein are protein libraries, wherein the protein library comprises antibodies. Provided herein are protein libraries, wherein the plurality of proteins are used to generate a peptidomimetic library. Provided herein are cell libraries comprising protein libraries as described herein.

[0009] Provided herein are vector libraries comprising a nucleic acid library described herein. Provided herein are cell libraries comprising a nucleic acid library described herein. Provided herein are cell libraries comprising a protein library described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0010] **Figure 1** depicts a schematic of G protein-coupled receptor (GPCR) ligand interaction surfaces.
- [0011] **Figure 2A** depicts a first schematic of an immunoglobulin scaffold.
- [0012] **Figure 2B** depicts a second schematic of an immunoglobulin scaffold.
- [0013] **Figure 3** depicts a schematic of a motif for placement in a scaffold.
- [0014] **Figure 4** depicts a schematic of a GPCR.
- [0015] **Figure 5** depicts schematics of segments for assembly of clonal fragments and non-clonal fragments.
- [0016] **Figure 6** depicts schematics of segments for assembly of clonal fragments and non-clonal fragments.
- [0017] **Figure 7** presents a diagram of steps demonstrating an exemplary process workflow for gene synthesis as disclosed herein.
- [0018] **Figure 8** illustrates an example of a computer system.
- [0019] **Figure 9** is a block diagram illustrating an architecture of a computer system.
- [0020] **Figure 10** is a diagram demonstrating a network configured to incorporate a plurality of computer systems, a plurality of cell phones and personal data assistants, and Network Attached Storage (NAS).
- [0021] **Figure 11** is a block diagram of a multiprocessor computer system using a shared virtual address memory space.
- [0022] **Figures 12A-12C** depict sequences of immunoglobulin scaffolds.
- [0023] **Figure 13** depicts sequences of G protein-coupled receptors scaffolds.
- [0024] **Figure 14** is a graph of normalized reads for a library for variable domain, heavy chains.
- [0025] **Figure 15** is a graph of normalized reads for a library for variable domain, light chains.
- [0026] **Figure 16** is a graph of normalized reads for a library for heavy chain complementarity determining region 3.
- [0027] **Figure 17A** is a plot of light chain frameworks assayed for folding.
- [0028] **Figure 17B** is a plot of light chain frameworks assayed for thermostability.
- [0029] **Figure 17C** is a plot of light chain frameworks assayed for motif display using FLAG tag.
- [0030] **Figure 17D** is a plot of light chain frameworks assayed for motif display using His tag.

- [0031] **Figure 18A** is a plot of heavy chain frameworks assayed for folding.
- [0032] **Figure 18B** is a plot of heavy chain frameworks assayed for stability.
- [0033] **Figure 18C** is a plot of heavy chain frameworks assayed for motif display using FLAG tag.
- [0034] **Figure 18D** is a plot of heavy chain frameworks assayed for motif display using His tag.
- [0035] **Figure 18E** is a plot of heavy chain frameworks assayed for expression.
- [0036] **Figure 18F** is a plot of heavy chain frameworks assayed for selection specificity.
- [0037] **Figures 19A-19C** depict images of G protein-coupled receptors visualized by fluorescent antibodies.
- [0038] **Figures 20A-20C** depict images of G protein-coupled receptors visualized by auto-fluorescent proteins.
- [0039] **Figure 21A** depicts a schematic of an immunoglobulin scaffold comprising a VH domain attached to a VL domain using a linker.
- [0040] **Figure 21B** depicts a schematic of a full-domain architecture of an immunoglobulin scaffold comprising a VH domain attached to a VL domain using a linker, a leader sequence, and pIII sequence.
- [0041] **Figure 21C** depicts a schematic of four framework elements (FW1, FW2, FW3, FW4) and the variable 3 CDR (L1, L2, L3) elements for a VL or VH domain.

DETAILED DESCRIPTION

[0042] The present disclosure employs, unless otherwise indicated, conventional molecular biology techniques, which are within the skill of the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art.

[0043] Definitions

[0044] Throughout this disclosure, various embodiments are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of any embodiments. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range to the tenth of the unit of the lower limit unless the context clearly dictates otherwise. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual values within that range, for example, 1.1, 2, 2.3, 5, and 5.9. This applies regardless

of the breadth of the range. The upper and lower limits of these intervening ranges may independently be included in the smaller ranges, and are also encompassed within the disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the disclosure, unless the context clearly dictates otherwise.

[0045] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of any embodiment. As used herein, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0046] Unless specifically stated or obvious from context, as used herein, the term “about” in reference to a number or range of numbers is understood to mean the stated number and numbers +/- 10% thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range.

[0047] Unless specifically stated, as used herein, the term “nucleic acid” encompasses double- or triple-stranded nucleic acids, as well as single-stranded molecules. In double- or triple-stranded nucleic acids, the nucleic acid strands need not be coextensive (i.e., a double-stranded nucleic acid need not be double-stranded along the entire length of both strands). Nucleic acid sequences, when provided, are listed in the 5' to 3' direction, unless stated otherwise. Methods described herein provide for the generation of isolated nucleic acids. Methods described herein additionally provide for the generation of isolated and purified nucleic acids. A “nucleic acid” as referred to herein can comprise at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, or more bases in length. Moreover, provided herein are methods for the synthesis of any number of polypeptide-segments encoding nucleotide sequences, including sequences encoding non-ribosomal peptides (NRPs), sequences encoding non-ribosomal peptide-synthetase (NRPS) modules and synthetic variants, polypeptide segments of other modular proteins, such as antibodies, polypeptide segments from other protein families, including non-coding DNA or RNA, such as regulatory sequences e.g. promoters, transcription factors, enhancers, siRNA,

shRNA, RNAi, miRNA, small nucleolar RNA derived from microRNA, or any functional or structural DNA or RNA unit of interest. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, intergenic DNA, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), small nucleolar RNA, ribozymes, complementary DNA (cDNA), which is a DNA representation of mRNA, usually obtained by reverse transcription of messenger RNA (mRNA) or by amplification; DNA molecules produced synthetically or by amplification, genomic DNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. cDNA encoding for a gene or gene fragment referred herein may comprise at least one region encoding for exon sequences without an intervening intron sequence in the genomic equivalent sequence.

[0048] GPCR Libraries

[0049] Provided herein are methods and compositions relating to G protein-coupled receptor (GPCR) binding libraries comprising nucleic acids encoding for a scaffold comprising a GPCR binding domain. Scaffolds as described herein can stably support a GPCR binding domain. The GPCR binding domain may be designed based on surface interactions of a GPCR ligand and the GPCR. Libraries as described herein may be further variegated to provide for variant libraries comprising nucleic acids each encoding for a predetermined variant of at least one predetermined reference nucleic acid sequence. Further described herein are protein libraries that may be generated when the nucleic acid libraries are translated. In some instances, nucleic acid libraries as described herein are transferred into cells to generate a cell library. Also provided herein are downstream applications for the libraries synthesized using methods described herein. Downstream applications include identification of variant nucleic acids or protein sequences with enhanced biologically relevant functions, e.g., improved stability, affinity, binding, functional activity, and for the treatment or prevention of a disease state associated with GPCR signaling.

[0050] Scaffold Libraries

[0051] Provided herein are libraries comprising nucleic acids encoding for a scaffold, wherein sequences for GPCR binding domains are placed in the scaffold. Scaffold described herein allow for improved stability for a range of GPCR binding domain encoding sequences when inserted into the scaffold, as compared to an unmodified scaffold. Exemplary scaffolds include, but are not limited to, a protein, a peptide, an immunoglobulin, derivatives thereof, or combinations thereof. In some instances, the scaffold is an immunoglobulin. Scaffolds as

described herein comprise improved functional activity, structural stability, expression, specificity, or a combination thereof. In some instances, scaffolds comprise long regions for supporting a GPCR binding domain.

[0052] Provided herein are libraries comprising nucleic acids encoding for a scaffold, wherein the scaffold is an immunoglobulin. In some instances, the immunoglobulin is an antibody. As used herein, the term antibody will be understood to include proteins having the characteristic two-armed, Y-shape of a typical antibody molecule as well as one or more fragments of an antibody that retain the ability to specifically bind to an antigen. Exemplary antibodies include, but are not limited to, a monoclonal antibody, a polyclonal antibody, a bi-specific antibody, a multispecific antibody, a grafted antibody, a human antibody, a humanized antibody, a synthetic antibody, a chimeric antibody, a camelized antibody, a single-chain Fvs (scFv) (including fragments in which the VL and VH are joined using recombinant methods by a synthetic or natural linker that enables them to be made as a single protein chain in which the VL and VH regions pair to form monovalent molecules, including single chain Fab and scFab), a single chain antibody, a Fab fragment (including monovalent fragments comprising the VL, VH, CL, and CH1 domains), a F(ab')₂ fragment (including bivalent fragments comprising two Fab fragments linked by a disulfide bridge at the hinge region), a Fd fragment (including fragments comprising the VH and CH1 fragment), a Fv fragment (including fragments comprising the VL and VH domains of a single arm of an antibody), a single-domain antibody (dAb or sdAb) (including fragments comprising a VH domain), an isolated complementarity determining region (CDR), a diabody (including fragments comprising bivalent dimers such as two VL and VH domains bound to each other and recognizing two different antigens), a fragment comprised of only a single monomeric variable domain, disulfide-linked Fvs (sdFv), an intrabody, an anti-idiotypic (anti-Id) antibody, or ab antigen-binding fragments thereof. In some instances, the libraries disclosed herein comprise nucleic acids encoding for a scaffold, wherein the scaffold is a Fv antibody, including Fv antibodies comprised of the minimum antibody fragment which contains a complete antigen-recognition and antigen-binding site. In some embodiments, the Fv antibody consists of a dimer of one heavy chain and one light chain variable domain in tight, non-covalent association, and the three hypervariable regions of each variable domain interact to define an antigen-binding site on the surface of the VH-VL dimer. In some embodiments, the six hypervariable regions confer antigen-binding specificity to the antibody. In some embodiments, a single variable domain (or half of an Fv comprising only three hypervariable regions specific for an antigen, including single domain antibodies isolated from camelid animals comprising one heavy chain variable domain such as VHH antibodies or nanobodies) has the ability to recognize

and bind antigen. In some instances, the libraries disclosed herein comprise nucleic acids encoding for a scaffold, wherein the scaffold is a single-chain Fv or scFv, including antibody fragments comprising a VH, a VL, or both a VH and VL domain, wherein both domains are present in a single polypeptide chain. In some embodiments, the Fv polypeptide further comprises a polypeptide linker between the VH and VL domains allowing the scFv to form the desired structure for antigen binding. In some instances, a scFv is linked to the Fc fragment or a VHH is linked to the Fc fragment (including minibodies). In some instances, the antibody comprises immunoglobulin molecules and immunologically active fragments of immunoglobulin molecules, e.g., molecules that contain an antigen binding site. Immunoglobulin molecules are of any type (e.g., IgG, IgE, IgM, IgD, IgA and IgY), class (e.g., IgG 1, IgG 2, IgG 3, IgG 4, IgA 1 and IgA 2) or subclass.

[0053] Libraries described herein comprising nucleic acids encoding for a scaffold, wherein the scaffold is an immunoglobulin, comprise variations in at least one region of the immunoglobulin. Exemplary regions of the antibody for variation include, but are not limited to, a complementarity-determining region (CDR), a variable domain, or a constant domain. In some instances, the CDR is CDR1, CDR2, or CDR3. In some instances, the CDR is a heavy domain including, but not limited to, CDR-H1, CDR-H2, and CDR-H3. In some instances, the CDR is a light domain including, but not limited to, CDR-L1, CDR-L2, and CDR-L3. In some instances, the variable domain is variable domain, light chain (VL) or variable domain, heavy chain (VH). In some instances, the VL domain comprises kappa or lambda chains. In some instances, the constant domain is constant domain, light chain (CL) or constant domain, heavy chain (CH).

[0054] Methods described herein provide for synthesis of libraries comprising nucleic acids encoding for a scaffold, wherein each nucleic acid encodes for a predetermined variant of at least one predetermined reference nucleic acid sequence. In some cases, the predetermined reference sequence is a nucleic acid sequence encoding for a protein, and the variant library comprises sequences encoding for variation of at least a single codon such that a plurality of different variants of a single residue in the subsequent protein encoded by the synthesized nucleic acid are generated by standard translation processes. In some instances, the scaffold library comprises varied nucleic acids collectively encoding variations at multiple positions. In some instances, the variant library comprises sequences encoding for variation of at least a single codon of a CDR-H1, CDR-H2, CDR-H3, CDR-L1, CDR-L2, CDR-L3, VL, or VH domain. In some instances, the variant library comprises sequences encoding for variation of multiple codons of a CDR-H1, CDR-H2, CDR-H3, CDR-L1, CDR-L2, CDR-L3, VL, or VH domain. In some instances, the variant library comprises sequences encoding for variation of multiple codons of framework

element 1 (FW1), framework element 2 (FW2), framework element 3 (FW3), or framework element 4 (FW4). An exemplary number of codons for variation include, but are not limited to, at least or about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 225, 250, 275, 300, or more than 300 codons.

[0055] In some instances, the at least one region of the immunoglobulin for variation is from heavy chain V-gene family, heavy chain D-gene family, heavy chain J-gene family, light chain V-gene family, or light chain J-gene family. In some instances, the light chain V-gene family comprises immunoglobulin kappa (IGK) gene or immunoglobulin lambda (IGL). Exemplary genes include, but are not limited to, IGHV1-18, IGHV1-69, IGHV1-8, IGHV3-21, IGHV3-23, IGHV3-30/33m, IGHV3-28, IGHV1-69, IGHV3-74, IGHV4-39, IGHV4-59/61, IGKV1-39, IGKV1-9, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV4-1, IGLV1-51, and IGLV2-14.

[0056] Provided herein are libraries comprising nucleic acids encoding for immunoglobulin scaffolds, wherein the libraries are synthesized with various numbers of fragments. In some instances, the fragments comprise the CDR-H1, CDR-H2, CDR-H3, CDR-L1, CDR-L2, CDR-L3, VL, or VH domain. In some instances, the fragments comprise framework element 1 (FW1), framework element 2 (FW2), framework element 3 (FW3), or framework element 4 (FW4). In some instances, the scaffold libraries are synthesized with at least or about 2 fragments, 3 fragments, 4 fragments, 5 fragments, or more than 5 fragments. The length of each of the nucleic acid fragments or average length of the nucleic acids synthesized may be at least or about 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, or more than 600 base pairs. In some instances, the length is about 50 to 600, 75 to 575, 100 to 550, 125 to 525, 150 to 500, 175 to 475, 200 to 450, 225 to 425, 250 to 400, 275 to 375, or 300 to 350 base pairs.

[0057] Libraries comprising nucleic acids encoding for immunoglobulin scaffolds as described herein comprise various lengths of amino acids when translated. In some instances, the length of each of the amino acid fragments or average length of the amino acid synthesized may be at least or about 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, or more than 150 amino acids. In some instances, the length of the amino acid is about 15 to 150, 20 to 145, 25 to 140, 30 to 135, 35 to 130, 40 to 125, 45 to 120, 50 to 115, 55 to 110, 60 to 110, 65 to 105, 70 to 100, or 75 to 95 amino acids. In some instances, the length of the amino acid is about 22 amino acids to about 75 amino acids. In some instances, the immunoglobulin scaffolds comprise at least or about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, or more than 5000 amino acids.

[0058] A number of variant sequences for the at least one region of the immunoglobulin for variation are de novo synthesized using methods as described herein. In some instances, a number of variant sequences is de novo synthesized for CDR-H1, CDR-H2, CDR-H3, CDR-L1, CDR-L2, CDR-L3, VL, VH, or combinations thereof. In some instances, a number of variant sequences is de novo synthesized for framework element 1 (FW1), framework element 2 (FW2), framework element 3 (FW3), or framework element 4 (FW4). The number of variant sequences may be at least or about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, or more than 500 sequences. In some instances, the number of variant sequences is at least or about 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, or more than 8000 sequences. In some instances, the number of variant sequences is about 10 to 500, 25 to 475, 50 to 450, 75 to 425, 100 to 400, 125 to 375, 150 to 350, 175 to 325, 200 to 300, 225 to 375, 250 to 350, or 275 to 325 sequences.

[0059] Variant sequences for the at least one region of the immunoglobulin, in some instances, vary in length or sequence. In some instances, the at least one region that is de novo synthesized is for CDR-H1, CDR-H2, CDR-H3, CDR-L1, CDR-L2, CDR-L3, VL, VH, or combinations thereof. In some instances, the at least one region that is de novo synthesized is for framework element 1 (FW1), framework element 2 (FW2), framework element 3 (FW3), or framework element 4 (FW4). In some instances, the variant sequence comprises at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more than 50 variant nucleotides or amino acids as compared to wild-type. In some instances, the variant sequence comprises at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 additional nucleotides or amino acids as compared to wild-type. In some instances, the variant sequence comprises at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 less nucleotides or amino acids as compared to wild-type. In some instances, the libraries comprise at least or about 10^1 , 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , or more than 10^{10} variants.

[0060] Following synthesis of scaffold libraries, scaffold libraries may be used for screening and analysis. For example, scaffold libraries are assayed for library displayability and panning. In some instances, displayability is assayed using a selectable tag. Exemplary tags include, but are not limited to, a radioactive label, a fluorescent label, an enzyme, a chemiluminescent tag, a colorimetric tag, an affinity tag or other labels or tags that are known in the art. In some instances, the tag is histidine, polyhistidine, myc, hemagglutinin (HA), or FLAG. In some instances, scaffold libraries are assayed by sequencing using various methods including, but not limited to, single-molecule real-time (SMRT) sequencing, Polony sequencing, sequencing by

ligation, reversible terminator sequencing, proton detection sequencing, ion semiconductor sequencing, nanopore sequencing, electronic sequencing, pyrosequencing, Maxam-Gilbert sequencing, chain termination (e.g., Sanger) sequencing, +S sequencing, or sequencing by synthesis.

[0061] In some instances, the scaffold libraries are assayed for functional activity, structural stability (e.g., thermal stable or pH stable), expression, specificity, or a combination thereof. In some instances, the scaffold libraries are assayed for scaffolds capable of folding. In some instances, a region of the antibody is assayed for functional activity, structural stability, expression, specificity, folding, or a combination thereof. For example, a VH region or VL region is assayed for functional activity, structural stability, expression, specificity, folding, or a combination thereof.

[0062] *GPCR Libraries*

[0063] Provided herein are G protein-coupled receptor (GPCR) binding libraries comprising nucleic acids encoding for scaffolds comprising sequences for GPCR binding domains. In some instances, the scaffolds are immunoglobulins. In some instances, the scaffolds comprising sequences for GPCR binding domains are determined by interactions between the GPCR binding domains and the GPCRs.

[0064] Provided herein are libraries comprising nucleic acids encoding scaffolds comprising GPCR binding domains, wherein the GPCR binding domains are designed based on surface interactions on the GPCRs. Exemplary GPCRs are seen in **Table 1**. In some instances, the GPCR binding domains interact with the amino- or carboxy-terminus of the GPCR. In some instances, the GPCR binding domains interact with at least one transmembrane domain including, but not limited to, transmembrane domain 1 (TM1), transmembrane domain 2 (TM2), transmembrane domain 3 (TM3), transmembrane domain 4 (TM4), transmembrane domain 5 (TM5), transmembrane domain 6 (TM6), and transmembrane domain 7 (TM7). In some instances, the GPCR binding domains interact with an intracellular surface of the GPCR. For example, the GPCR binding domains interact with at least one intracellular loop including, but not limited to, intracellular loop 1 (ICL1), intracellular loop 2 (ICL2), and intracellular loop 3 (ICL3). In some instances, the GPCR binding domains interact with an extracellular surface of the GPCR. *See FIG. 1*. For example, the GPCR binding domains interact with at least one extracellular domain (ECD) or extracellular loop (ECL) of the GPCR. The extracellular loops include, but are not limited to, extracellular loop 1 (ECL1), extracellular loop 2 (ECL2), and extracellular loop 3 (ECL3).

Table 1. List of GPCRs

GPCR	Gene Name	Accession Number
5-hydroxytryptamine receptor 1A	HTR1A	P08908
5-hydroxytryptamine receptor 1B	HTR1B	P28222
5-hydroxytryptamine receptor 1D	HTR1D	P28221
5-hydroxytryptamine receptor 1E	HTR1E	P28566
5-hydroxytryptamine receptor 1F	HTR1F	P30939
5-hydroxytryptamine receptor 2A	HTR2A	P28223
5-hydroxytryptamine receptor 2B	HTR2B	P41595
5-hydroxytryptamine receptor 2C	HTR2C	P28335
5-hydroxytryptamine receptor 4	HTR4	Q13639
5-hydroxytryptamine receptor 5A	HTR5A	P47898
5-hydroxytryptamine receptor 6	HTR6	P50406
5-hydroxytryptamine receptor	HTR7	P34969
Adenosine receptor A1	ADORA1	P30542
Adenosine receptor A2a	ADORA2A	P29274
Adenosine receptor A2b	ADORA2B	P29275
Adenosine A3 receptor	ADORA3	P33765
Muscarinic acetylcholine receptor M1	CHRM1	P11229
Muscarinic acetylcholine receptor M2	CHRM2	P08172
Muscarinic acetylcholine receptor M3	CHRM3	P20309
Muscarinic acetylcholine receptor M4	CHRM4	P08173
Muscarinic acetylcholine receptor M5	CHRM5	P08912
Adrenocorticotrophic hormone receptor	MC2R	Q01718
α -1A adrenergic receptor	ADRA1A	P35348
α -1B adrenergic receptor	ADRA1B	P35368
α -1D adrenergic receptor	ADRA1D	P25100
α -2A adrenergic receptor	ADRA2A	P08913
α -2B adrenergic receptor	ADRA2B	P18089
α -2C adrenergic receptor	ADRA2C	P18825
β -1 adrenergic receptor	ADRB1	P08588
β -2 adrenergic receptor	ADRB2	P07550
β -3 adrenergic receptor	ADRB3	P13945
Type-1 angiotensin II receptor	AGTR1	P30556
Duffy antigen/chemokine receptor	DARC	Q16570
Endothelin-1 receptor	EDNRA	P25101
Endothelin B receptor	EDNRB	P24530
<i>N</i> -formyl peptide receptor 2	FPR2	P25090
Follicle-stimulating hormone receptor	FSHR	P23945
Galanin receptor type 1	GALR1	P47211
Galanin receptor type 2	GALR2	O43603
Galanin receptor type 3	GALR3	O60755
Gastrin/cholecystokinin type B receptor	CCKBR	P32239
Gonadotropin-releasing hormone receptor	GNRHR	P30968
Putative gonadotropin-releasing hormone II receptor	GNRHR2	Q96P88

G-protein coupled oestrogen receptor 1	GPOR	Q99527
Uracil nucleotide/cysteinyl leukotriene receptor	GPR17	Q13304
Putative G-protein coupled receptor 44	GPR44	Q9Y5Y4
G-protein coupled receptor 55	GPR55	Q9Y2T6
Gastrin-releasing peptide receptor	GRPR	P30550
Histamine H1 receptor	HRH1	P35367
Histamine H2 receptor	HRH2	P25021
Histamine H3 receptor	HRH3	Q9Y5N1
Histamine H4 receptor	HRH4	Q9H3N8
KISS-1 receptor	KISS1R	Q969F8
Lysophosphatidic acid receptor 1	LPAR1	Q92633
Lysophosphatidic acid receptor 2	LPAR2	Q9HBW0
Lysophosphatidic acid receptor 3	LPAR3	Q9UBY5
Lysophosphatidic acid receptor 4	LPAR4	Q99677
Lysophosphatidic acid receptor 6	LPAR6	P43657
Lutropin-choriogonadotropic hormone receptor	LHCGR	P22888
Leukotriene B4 receptor 1	LTB4R	Q15722
Leukotriene B4 receptor 2	LTB4R2	Q9NPC1
Melanocortin receptor 3	MC3R	P41968
Melanocortin receptor 4	MC4R	P32245
Melanocortin receptor 5	MC5R	P33032
Olfactory receptor 10G9	OR10G9	Q8NGN4
Olfactory receptor 10H1	OR10H1	Q9Y4A9
Olfactory receptor 10H2	OR10H2	Q60403
Olfactory receptor 10H3	OR10H3	Q60404
Olfactory receptor 10H4	OR10H4	Q8NGA5
Olfactory receptor 10H5	OR10H5	Q8NGA6
Olfactory receptor 10J1	OR10J1	P30954
Olfactory receptor 10J3	OR10J3	Q5JRS4
Olfactory receptor 10J5	OR10J5	Q8NHC4
Olfactory receptor 10K1	OR10K1	Q8NGX5
Olfactory receptor 10K2	OR10K2	Q6IF99
Olfactory receptor 10P1	OR10P1	Q8NGE3
Olfactory receptor 10Q1	OR10Q1	Q8NGQ4
Olfactory receptor 10R2	OR10R2	Q8NGX6
Olfactory receptor 10S1	OR10S1	Q8NGN2
Olfactory receptor 10T2	OR10T2	Q8NGX3
Olfactory receptor 10V1	OR10V1	Q8NGI7
Olfactory receptor 10W1	OR10W1	Q8NGF6
Olfactory receptor 14A2	OR14A2	Q96R54
Olfactory receptor 14C36	OR14C36	Q8NHC7
Olfactory receptor 14I1	OR14I1	A6ND48
Olfactory receptor 14J1	OR14J1	Q9UGF5
Olfactory receptor 14K1	OR14K1	Q8NGZ2
Olfactory receptor 2A12	OR2A12	Q8NGT7
Olfactory receptor 2A14	OR2A14	Q96R47
Olfactory receptor 2A25	OR2A25	A4D2G3
Olfactory receptor 2AG1	OR2AG1	Q9H205

Olfactory receptor 2AG2	OR2AG2	A6NM03
Olfactory receptor 2AJ1	OR2AJ1	Q8NGZ0
Olfactory receptor 2AK2	OR2AK2	Q8NG84
Olfactory receptor 2AP1	OR2AP1	Q8NGE2
Olfactory receptor 2AT4	OR2AT4	A6NND4
Olfactory receptor 51I2	OR51I2	Q9H344
Olfactory receptor 51J1	OR51J1	Q9H342
Olfactory receptor 51L1	OR51L1	Q8NGJ5
Olfactory receptor 51M1	OR51M1	Q9H341
Olfactory receptor 51Q1	OR51Q1	Q8NH59
Olfactory receptor 51S1	OR51S1	Q8NGJ8
Olfactory receptor 51T1	OR51T1	Q8NGJ9
Olfactory receptor 51V1	OR51V1	Q9H2C8
Olfactory receptor 52A1	OR52A1	Q9UKL2
Olfactory receptor 52A5	OR52A5	Q9H2C5
Olfactory receptor 52B2	OR52B2	Q96RD2
Olfactory receptor 52B4	OR52B4	Q8NGK2
Olfactory receptor 52B6	OR52B6	Q8NGF0
Olfactory receptor 52D1	OR52D1	Q9H346
Olfactory receptor 52E2	OR52E2	Q8NGJ4
Olfactory receptor 52E4	OR52E4	Q8NGH9
Olfactory receptor 52E5	OR52E5	Q8NH55
Olfactory receptor 52E6	OR52E6	Q96RD3
Olfactory receptor 52E8	OR52E8	Q6IFG1
Olfactory receptor 52H1	OR52H1	Q8NGJ2
Olfactory receptor 52I1	OR52I1	Q8NGK6
Olfactory receptor 52I2	OR52I2	Q8NH67
Olfactory receptor 52K1	OR52K1	Q8NGK4
Olfactory receptor 52K2	OR52K2	Q8NGK3
Olfactory receptor 52L1	OR52L1	Q8NGH7
Olfactory receptor 52M1	OR52M1	Q8NGK5
Olfactory receptor 52N1	OR52N1	Q8NH53
Olfactory receptor 52N2	OR52N2	Q8NGI0
Olfactory receptor 52N4	OR52N4	Q8NGI2
Olfactory receptor 52N5	OR52N5	Q8NH56
Olfactory receptor 52R1	OR52R1	Q8NGF1
Olfactory receptor 52W1	OR52W1	Q6IF63
Red-sensitive opsin	OPN1LW	P04000
Visual pigment-like receptor peropsin	RRH	O14718
Olfactory receptor 1A1	OR1A1	Q9P1Q5
Olfactory receptor 1A2	OR1A2	Q9Y585
Olfactory receptor 1B1	OR1B1	Q8NGR6
Olfactory receptor 1C1	OR1C1	Q15619
Olfactory receptor 1D2	OR1D2	P34982
Olfactory receptor 1F1	OR1F1	O43749
Olfactory receptor 1F12	OR1F12	Q8NHA8
Olfactory receptor 1G1	OR1G1	P47890
Olfactory receptor 1I1	OR1I1	O60431

Olfactory receptor 1J1	OR1J1	Q8NGS3
Olfactory receptor 1J2	OR1J2	Q8NGS2
Olfactory receptor 1J4	OR1J4	Q8NGS1
Olfactory receptor 1K1	OR1K1	Q8NGR3
Olfactory receptor 1L1	OR1L1	Q8NH94
Olfactory receptor 1L3	OR1L3	Q8NH93
Olfactory receptor 1L4	OR1L4	Q8NGR5
Olfactory receptor 1L6	OR1L6	Q8NGR2
Olfactory receptor 1L8	OR1L8	Q8NGR8
Olfactory receptor 1M1	OR1M1	Q8NGA1
Olfactory receptor 1N1	OR1N1	Q8NGS0
Olfactory receptor 1N2	OR1N2	Q8NGR9
Olfactory receptor 1Q1	OR1Q1	Q15612
Olfactory receptor 1S1	OR1S1	Q8NH92
Olfactory receptor 1S2	OR1S2	Q8NGQ3
Olfactory receptor 2A2	OR2A2	Q6IF42
Olfactory receptor 2A4	OR2A4	Q95047
Olfactory receptor 2B2	OR2B2	Q9GZK3
Putative olfactory receptor 2B3	OR2B3	Q76000
Olfactory receptor 2B6	OR2B6	P58173
Putative olfactory receptor 2B8	OR2B8P	P59922
Olfactory receptor 2T5	OR2T5	Q6IEZ7
Olfactory receptor 2T6	OR2T6	Q8NHC8
Olfactory receptor 2T8	OR2T8	A6NH00
Olfactory receptor 2V1	OR2V1	Q8NHB1
Olfactory receptor 2V2	OR2V2	Q96R30
Olfactory receptor 2W1	OR2W1	Q9Y3N9
Olfactory receptor 2W3	OR2W3	Q7Z3T1
Olfactory receptor 2Y1	OR2Y1	Q8NGV0
Olfactory receptor 2Z1	OR2Z1	Q8NG97
Olfactory receptor 3A1	OR3A1	P47881
Olfactory receptor 3A2	OR3A2	P47893
Olfactory receptor 3A3	OR3A3	P47888
Olfactory receptor 3A4	OR3A4	P47883
Putative olfactory receptor 4A4	OR4A4P	Q8NGN8
Olfactory receptor 4A5	OR4A5	Q8NH83
Olfactory receptor 4A8	OR4A8P	P0C604
Olfactory receptor 4B1	OR4B1	Q8NGF8
Olfactory receptor 4C3	OR4C3	Q8NH37
Olfactory receptor 4C5	OR4C5	Q8NGB2
Olfactory receptor 4C6	OR4C6	Q8NH72
Olfactory receptor 4C11	OR4C11	Q6IEV9
Olfactory receptor 4C12	OR4C12	Q96R67
Olfactory receptor 4C13	OR4C13	Q8NGP0
Olfactory receptor 4C15	OR4C15	Q8NGM1
Olfactory receptor 4C16	OR4C16	Q8NGL9
Olfactory receptor 4D1	OR4D1	Q15615
Olfactory receptor 4D2	OR4D2	P58180

Olfactory receptor 4D5	OR4D5	Q8NGN0
Olfactory receptor 4D6	OR4D6	Q8NGJ1
Olfactory receptor 4D9	OR4D9	Q8NGE8
Olfactory receptor 4D10	OR4D10	Q8NGI6
Olfactory receptor 4D11	OR4D11	Q8NGI4
Olfactory receptor 5B17	OR5B17	Q8NGF7
Olfactory receptor 5B21	OR5B21	A6NL26
Olfactory receptor 5C1	OR5C1	Q8NGR4
Olfactory receptor 5D13	OR5D13	Q8NGL4
Olfactory receptor 5D14	OR5D14	Q8NGL3
Olfactory receptor 5D16	OR5D16	Q8NGK9
Olfactory receptor 5D18	OR5D18	Q8NGL1
Olfactory receptor 5F1	OR5F1	O95221
Olfactory receptor 5H1	OR5H1	A6NKK0
Olfactory receptor 5H2	OR5H2	Q8NGV7
Olfactory receptor 5H6	OR5H6	Q8NGV6
Olfactory receptor 5I1	OR5I1	Q13606
Olfactory receptor 5J2	OR5J2	Q8NH18
Olfactory receptor 5K1	OR5K1	Q8NHB7
Olfactory receptor 5K2	OR5K2	Q8NHB8
Olfactory receptor 5K3	OR5K3	A6NET4
Olfactory receptor 5K4	OR5K4	A6NMS3
Olfactory receptor 5L1	OR5L1	Q8NGL2
Olfactory receptor 5L2	OR5L2	Q8NGL0
Olfactory receptor 5M1	OR5M1	Q8NGP8
Olfactory receptor 5M3	OR5M3	Q8NGP4
Olfactory receptor 5M8	OR5M8	Q8NGP6
Olfactory receptor 5M9	OR5M9	Q8NGP3
Olfactory receptor 5M10	OR5M10	Q6IEU7
Olfactory receptor 5M11	OR5M11	Q96RB7
Olfactory receptor 5P2	OR5P2	Q8WZ92
Olfactory receptor 5P3	OR5P3	Q8WZ94
Olfactory receptor 5R1	OR5R1	Q8NH85
Olfactory receptor 5T1	OR5T1	Q8NG75
Olfactory receptor 5T2	OR5T2	Q8NGG2
Olfactory receptor 5T3	OR5T3	Q8NGG3
Olfactory receptor 5W2	OR5W2	Q8NH69
Olfactory receptor 7G3	OR7G3	Q8NG95
Olfactory receptor 8A1	OR8A1	Q8NGG7
Olfactory receptor 8B3	OR8B3	Q8NGG8
Olfactory receptor 8B4	OR8B4	Q96RC9
Olfactory receptor 8B8	OR8B8	Q15620
Olfactory receptor 8B12	OR8B12	Q8NGG6
Olfactory receptor 8D1	OR8D1	Q8WZ84
Olfactory receptor 8D2	OR8D2	Q9GZM6
Olfactory receptor 8D4	OR8D4	Q8NGM9
Orexin receptor type 1	HCRTR1	O43613
Orexin receptor type 2	HCRTR2	O43614

Oxoecicosanoid receptor 1	OXER1	Q8TDS5
Oxytocin receptor	OXTR	P30559
P2Y purinoceptor 1	P2RY1	P47900
P2Y purinoceptor 2	P2RY2	P41231
P2Y purinoceptor 4	P2RY4	P51582
P2Y purinoceptor 6	P2RY6	Q15077
P2Y purinoceptor 8	P2RY8	Q86VZ1
Putative P2Y purinoceptor 10	P2RY10	O00398
P2Y purinoceptor 11	P2RY11	Q96G91
P2Y purinoceptor 12	P2RY12	Q9H244
P2Y purinoceptor 13	P2RY13	Q9BPV8
P2Y purinoceptor 14	P2RY14	Q15391
Proteinase-activated receptor 1	F2R	P25116
Proteinase-activated receptor 2	F2RL1	P55085
Proteinase-activated receptor 3	F2RL2	O00254
Proteinase-activated receptor 4	F2RL3	Q96RI0
Prostaglandin D2 receptor	PTGDR	Q13258
Prostaglandin E2 receptor EP1 subtype	PTGER1	P34995
Prostaglandin E2 receptor EP2 subtype	PTGER2	P43116
Prostaglandin E2 receptor EP3 subtype	PTGER3	P43115
Prostaglandin E2 receptor EP4 subtype	PTGER4	P35408
Type-2 angiotensin II receptor (AT2)	AGTR2	P50052
Apelin receptor	APLNR	P35414
B1 bradykinin receptor	BDKRB1	P46663
B2 bradykinin receptor	BDKRB2	P30411
C5a anaphylatoxin chemotactic receptor	C5AR1	P21730
Cholecystokinin receptor type A	CCKAR	P32238
C-C chemokine receptor type 10	CCR10	P46092
C-C chemokine receptor type 1	CCR1	P32246
C-C chemokine receptor type 2	CCR2	P41597
C-C chemokine receptor type 3	CCR3	P51677
C-C chemokine receptor type 4	CCR4	P51679
C-C chemokine receptor type 5	CCR5	P51681
C-C chemokine receptor type 6	CCR6	P51684
C-C chemokine receptor type 7	CCR7	P32248
C-C chemokine receptor type 8	CCR8	P51685
C-C chemokine receptor type 9	CCR9	P51686
Cysteinyl leukotriene receptor 1	CYSLTR1	Q9Y271
Cysteinyl leukotriene receptor 2	CYSLTR2	Q9NS75
Cannabinoid receptor 1	CNR1	P21554
Cannabinoid receptor 2	CNR2	P34972
CX3C chemokine receptor 1	CX3CR1	P49238
High affinity interleukin-8 receptor A	IL8RA	P25024
High affinity interleukin-8 receptor B	IL8RB	P25025
C-X-C chemokine receptor type 3	CXCR3	P49682
C-X-C chemokine receptor type 4	CXCR4	P61073
C-X-C chemokine receptor type 6	CXCR6	O00574
C-X-C chemokine receptor type 7	CXCR7	P25106

D(1A) dopamine receptor	DRD1	P21728
D(2) dopamine receptor	DRD2	P14416
D(3) dopamine receptor	DRD3	P35462
D(4) dopamine receptor	DRD4	P21917
D(1B) dopamine receptor	DRD5	P21918
Melanocyte-stimulating hormone receptor	MC1R	Q01726
Melatonin receptor type 1A	MTNR1A	P48039
Melatonin receptor type 1B	MTNR1B	P49286
Substance-P receptor	TACR1	P25103
Substance-K receptor	TACR2	P21452
Neuromedin-K receptor	TACR3	P29371
Neuromedin-B receptor	NMBR	P28336
Neuropeptides B/W receptor type 1	NPBWR1	P48145
Neuropeptides B/W receptor type 2	NPBWR2	P48146
Neuropeptide FF receptor 1	NPFFR1	Q9GZQ6
Neuropeptide FF receptor 2	NPFFR2	Q9Y5X5
Neuropeptide Y receptor type 1	NPY1R	P25929
Neuropeptide Y receptor type 2	NPY2R	P49146
Neuropeptide Y receptor type 4	PPYR1	P50391
Neuropeptide Y receptor type 5	NPY5R	Q15761
Neurotensin receptor type 1	NTSR1	P30989
Neurotensin receptor type 2	NTSR2	Q95665
Olfactory receptor 10A2	OR10A2	Q9H208
Olfactory receptor 10A3	OR10A3	P58181
Olfactory receptor 10A4	OR10A4	Q9H209
Olfactory receptor 10A5	OR10A5	Q9H207
Olfactory receptor 10A6	OR10A6	Q8NH74
Olfactory receptor 10A7	OR10A7	Q8NGE5
Olfactory receptor 10AD1	OR10AD1	Q8NGE0
Olfactory receptor 10AG1	OR10AG1	Q8NH19
Olfactory receptor 10C1	OR10C1	Q96KK4
Olfactory receptor 10G2	OR10G2	Q8NGC3
Olfactory receptor 10G3	OR10G3	Q8NGC4
Olfactory receptor 10G4	OR10G4	Q8NGN3
Olfactory receptor 10G6	OR10G6	Q8NH81
Olfactory receptor 10G7	OR10G7	Q8NGN6
Olfactory receptor 10G8	OR10G8	Q8NGN5
Olfactory receptor 2T10	OR2T10	Q8NGZ9
Olfactory receptor 2T11	OR2T11	Q8NH01
Olfactory receptor 2T12	OR2T12	Q8NG77
Olfactory receptor 2T27	OR2T27	Q8NH04
Olfactory receptor 2T29	OR2T29	Q8NH02
Olfactory receptor 2T33	OR2T33	Q8NG76
Olfactory receptor 2T34	OR2T34	Q8NGX1
Olfactory receptor 2T35	OR2T35	Q8NGX2
Olfactory receptor 4A15	OR4A15	Q8NGL6
Olfactory receptor 4A16	OR4A16	Q8NH70
Olfactory receptor 4A47	OR4A47	Q6IF82

Olfactory receptor 4C45	OR4C45	A6NMZ5
Olfactory receptor 4C46	OR4C46	A6NHA9
Olfactory receptor 4F15	OR4F15	Q8NGB8
Olfactory receptor 4F17	OR4F17	Q8NGA8
Olfactory receptor 4F21	OR4F21	O95013
Olfactory receptor 51A2	OR51A2	Q8NGJ7
Olfactory receptor 51A4	OR51A4	Q8NGJ6
Olfactory receptor 51A7	OR51A7	Q8NH64
Olfactory receptor 51B2	OR51B2	Q9Y5P1
Olfactory receptor 51B4	OR51B4	Q9Y5P0
Olfactory receptor 51B5	OR51B5	Q9H339
Olfactory receptor 51B5	OR51B6	Q9H340
Olfactory receptor 51D1	OR51D1	Q8NGF3
Olfactory receptor 51E1	OR51E1	Q8TCB6
Olfactory receptor 51E2	OR51E2	Q9H255
Olfactory receptor 51F1	OR51F1	A6NGY5
Olfactory receptor 51F2	OR51F2	Q8NH61
Olfactory receptor 51G1	OR51G1	Q8NGK1
Olfactory receptor 51G2	OR51G2	Q8NGK0
Putative olfactory receptor 51H1	OR51H1P	Q8NH63
Olfactory receptor 51I1	OR51I1	Q9H343
Olfactory receptor 56A1	OR56A1	Q8NGH5
Olfactory receptor 56A3	OR56A3	Q8NH54
Olfactory receptor 56A4	OR56A4	Q8NGH8
Olfactory receptor 56A5	OR56A5	P0C7T3
Olfactory receptor 56B1	OR56B1	Q8NGI3
Olfactory receptor 56B4	OR56B4	Q8NH76
Olfactory receptor 5AC2	OR5AC2	Q9NZP5
Olfactory receptor 5AK2	OR5AK2	Q8NH90
Olfactory receptor 5AN1	OR5AN1	Q8NGI8
Olfactory receptor 5AP2	OR5AP2	Q8NGF4
Olfactory receptor 5AR1	OR5AR1	Q8NGP9
Olfactory receptor 5AS1	OR5AS1	Q8N127
Olfactory receptor 5AU1	OR5AU1	Q8NGC0
Olfactory receptor 5H14	OR5H14	A6NHG9
Olfactory receptor 5H15	OR5H15	A6NDH6
Olfactory receptor 6C65	OR6C65	A6NJZ3
Olfactory receptor 6C68	OR6C68	A6NDL8
Olfactory receptor 6C70	OR6C70	A6NIJ9
Olfactory receptor 6C74	OR6C74	A6NCV1
Olfactory receptor 6C75	OR6C75	A6NL08
Olfactory receptor 6C76	OR6C76	A6NM76
Olfactory receptor 7E24	OR7E24	Q6IFN5
Opsin-3	OPN3	Q9H1Y3
Melanopsin	OPN4	Q9UHM6
Opsin-5	OPN5	Q6U736
δ -type opioid receptor	OPRD1	P41143
κ -type opioid receptor	OPRK1	P41145

μ-type opioid receptor	OPRM1	P35372
Nociceptin receptor	OPRL1	P41146
Blue-sensitive opsin	OPN1SW	P03999
Rhodopsin	RHO	P08100
Green-sensitive opsin	OPN1MW	P04001
Olfactory receptor 2B11	OR2B11	Q5JQS5
Olfactory receptor 2C1	OR2C1	O95371
Olfactory receptor 2C3	OR2C3	Q8N628
Olfactory receptor 2D2	OR2D2	Q9H210
Olfactory receptor 2D3	OR2D3	Q8NGH3
Olfactory receptor 2F1	OR2F1	Q13607
Olfactory receptor 2F2	OR2F2	O95006
Olfactory receptor 2G2	OR2G2	Q8NGZ5
Olfactory receptor 2G3	OR2G3	Q8NGZ4
Olfactory receptor 2G6	OR2G6	Q5TZ20
Olfactory receptor 2H1	OR2H1	Q9GZK4
Olfactory receptor 2H2	OR2H2	O95918
Putative olfactory receptor 2I1	OR2I1P	Q8NGU4
Olfactory receptor 2J1	OR2J1	Q9GZK6
Olfactory receptor 2J2	OR2J2	O76002
Olfactory receptor 2J3	OR2J3	O76001
Olfactory receptor 2K2	OR2K2	Q8NGT1
Olfactory receptor 2L2	OR2L2	Q8NH16
Olfactory receptor 2L3	OR2L3	Q8NG85
Olfactory receptor 2L5	OR2L5	Q8NG80
Olfactory receptor 2L8	OR2L8	Q8NGY9
Olfactory receptor 2L13	OR2L13	Q8N349
Olfactory receptor 2M2	OR2M2	Q96R28
Olfactory receptor 2M3	OR2M3	Q8NG83
Olfactory receptor 2M4	OR2M4	Q96R27
Olfactory receptor 2M5	OR2M5	A3KFT3
Olfactory receptor 2M7	OR2M7	Q8NG81
Olfactory receptor 2S2	OR2S2	Q9NQN1
Olfactory receptor 2T1	OR2T1	O43869
Olfactory receptor 2T2	OR2T2	Q6IF00
Olfactory receptor 2T3	OR2T3	Q8NH03
Olfactory receptor 2T4	OR2T4	Q8NH00
Olfactory receptor 4E1	OR4E1	P0C645
Olfactory receptor 4E2	OR4E2	Q8NGC2
Olfactory receptor 4F3/4F16/4F29	OR4F3	Q6IEY1
Olfactory receptor 4F4	OR4F4	Q96R69
Olfactory receptor 4F5	OR4F5	Q8NH21
Olfactory receptor 4F6	OR4F6	Q8NGB9
Olfactory receptor 4K1	OR4K1	Q8NGD4
Olfactory receptor 4K2	OR4K2	Q8NGD2
Olfactory receptor 4K3	OR4K3	Q96R72
Olfactory receptor 4K5	OR4K5	Q8NGD3
Olfactory receptor 4K13	OR4K13	Q8NH42

Olfactory receptor 4K14	OR4K14	Q8NGD5
Olfactory receptor 4K15	OR4K15	Q8NH41
Olfactory receptor 4K17	OR4K17	Q8NGC6
Olfactory receptor 4L1	OR4L1	Q8NH43
Olfactory receptor 4M1	OR4M1	Q8NGD0
Olfactory receptor 4M2	OR4M2	Q8NGB6
Olfactory receptor 4N2	OR4N2	Q8NGD1
Olfactory receptor 4N4	OR4N4	Q8N0Y3
Olfactory receptor 4N5	OR4N5	Q8IXE1
Olfactory receptor 4P4	OR4P4	Q8NGL7
Olfactory receptor 4Q2	OR4Q2	P0C623
Olfactory receptor 4Q3	OR4Q3	Q8NH05
Olfactory receptor 4S1	OR4S1	Q8NGB4
Olfactory receptor 4S2	OR4S2	Q8NH73
Olfactory receptor 4X1	OR4X1	Q8NH49
Olfactory receptor 4X2	OR4X2	Q8NGF9
Olfactory receptor 5A1	OR5A1	Q8NGJ0
Olfactory receptor 5A2	OR5A2	Q8NGI9
Olfactory receptor 5B2	OR5B2	Q96R09
Olfactory receptor 5B3	OR5B3	Q8NH48
Olfactory receptor 5B12	OR5B12	Q96R08
Olfactory receptor 6A2	OR6A2	O95222
Olfactory receptor 6B1	OR6B1	O95007
Olfactory receptor 6B2	OR6B2	Q6IFH4
Olfactory receptor 6B3	OR6B3	Q8NGW1
Olfactory receptor 6C1	OR6C1	Q96RD1
Olfactory receptor 6C2	OR6C2	Q9NZP2
Olfactory receptor 6C3	OR6C3	Q9NZP0
Olfactory receptor 6C4	OR6C4	Q8NGE1
Olfactory receptor 6C6	OR6C6	A6NF89
Olfactory receptor 6F1	OR6F1	Q8NGZ6
Olfactory receptor 6J1	OR6J1	Q8NGC5
Olfactory receptor 6K2	OR6K2	Q8NGY2
Olfactory receptor 6K3	OR6K3	Q8NGY3
Olfactory receptor 6K6	OR6K6	Q8NGW6
Olfactory receptor 6M1	OR6M1	Q8NGM8
Olfactory receptor 6N1	OR6N1	Q8NGY5
Olfactory receptor 6N2	OR6N2	Q8NGY6
Olfactory receptor 6P1	OR6P1	Q8NGX9
Olfactory receptor 6Q1	OR6Q1	Q8NGQ2
Olfactory receptor 6S1	OR6S1	Q8NH40
Olfactory receptor 6T1	OR6T1	Q8NGN1
Olfactory receptor 6V1	OR6V1	Q8N148
Olfactory receptor 6X1	OR6X1	Q8NH79
Olfactory receptor 6Y1	OR6Y1	Q8NGX8
Olfactory receptor 7A5	OR7A5	Q15622
Olfactory receptor 7A10	OR7A10	O76100
Olfactory receptor 7A17	OR7A17	O14581

Olfactory receptor 7C1	OR7C1	O76099
Olfactory receptor 7C2	OR7C2	O60412
Olfactory receptor 7D4	OR7D4	Q8NG98
Olfactory receptor 7G1	OR7G1	Q8NGA0
Olfactory receptor 7G2	OR7G2	Q8NG99
Prostaglandin F2- α receptor	PTGFR	P43088
Prostacyclin receptor	PTGIR	P43119
Prolactin-releasing peptide receptor	PRLHR	P49683
Platelet-activating factor receptor	PTAFR	P25105
Pyroglutamylated RFamide peptide receptor	QRFPR	Q96P65
RPE-retinal G protein-coupled receptor	RGR	P47804
Sphingosine 1-phosphate receptor 1	S1PR1	P21453
Sphingosine 1-phosphate receptor 2	S1PR2	O95136
Sphingosine 1-phosphate receptor 3	S1PR3	Q99500
Sphingosine 1-phosphate receptor 4	S1PR4	O95977
Sphingosine 1-phosphate receptor 5	S1PR5	Q9H228
Somatostatin receptor type 1	SSTR1	P30872
Somatostatin receptor type 2	SSTR2	P30874
Somatostatin receptor type 3	SSTR3	P32745
Somatostatin receptor type 4	SSTR4	P31391
Somatostatin receptor type 5	SSTR5	P35346
Thromboxane A2 receptor	TBXA2R	P21731
Trace amine-associated receptor 1	TAAR1	Q96RJ0
Trace amine-associated receptor 2	TAAR2	Q9P1P5
Putative trace amine-associated receptor 3	TAAR3	Q9P1P4
Trace amine-associated receptor 5	TAAR5	O14804
Trace amine-associated receptor 6	TAAR6	Q96RI8
Trace amine-associated receptor 8	TAAR8	Q969N4
Trace amine-associated receptor 9	TAAR9	Q96RI9
Thyrotropin receptor	TSHR	P16473
Vasopressin V1a receptor	AVPR1A	P37288
Vasopressin V1b receptor	AVPR1B	P47901
Vasopressin V2 receptor	AVPR2	P30518
Chemokine XC receptor 1	XCR1	P46094
Brain-specific angiogenesis inhibitor 1	BAI1	O14514
Brain-specific angiogenesis inhibitor 2	BAI2	O60241
Brain-specific angiogenesis inhibitor 3	BAI3	O60242
Calcitonin receptor	CALCR	P30988
Calcitonin gene-related peptide type 1 receptor	CALCRL	Q16602
Corticotropin-releasing factor receptor 1	CRHR1	P34998
Corticotropin-releasing factor receptor 2	CRHR2	Q13324
Growth hormone-releasing hormone receptor	GHRHR	Q02643
Gastric inhibitory polypeptide receptor	GIPR	P48546
Glucagon-like peptide 1 receptor	GLP1R	P43220
Glucagon-like peptide 2 receptor	GLP2R	O95838
Glucagon receptor	GCGR	P47871
Pituitary adenylate cyclase-activating polypeptide type I receptor	ADCYAP1R1	P41586

Taste receptor 1 member 2	TAS1R2	Q8TE23
Parathyroid hormone receptor 1	PTH1R	Q03431
Parathyroid hormone 2 receptor	PTH2R	P49190
Secretin receptor	SCTR	P47872
Vasoactive intestinal polypeptide receptor 1	VIPR1	P32241
Vasoactive intestinal polypeptide receptor 2	VIPR2	P41587
Frizzled-10	FZD10	Q9ULW2
Frizzled-1	FZD1	Q9UP38
Frizzled-2	FZD2	Q14332
Frizzled-3	FZD3	Q9NPG1
Frizzled-4	FZD4	Q9ULV1
Frizzled-5	FZD5	Q13467
Frizzled-6	FZD6	O60353
Frizzled-7	FZD7	O75084
Frizzled-8	FZD8	Q9H461
Frizzled-9	FZD9 (FZD3)	O00144
Smoothened homologue	SMO (SMOH)	Q99835
Extracellular calcium-sensing receptor	CASR	P41180
GABA type B receptor 1 subunit 1	GABBR1	Q9UBS5
GABA type B receptor subunit 2	GABBR2	O75899
GPCR family C group 6 member A	GPRC6A	Q5T6X5
Metabotropic glutamate receptor 1	GRM1	Q13255
Metabotropic glutamate receptor 2	GRM2	Q14416
Metabotropic glutamate receptor 3	GRM3	Q14832
Metabotropic glutamate receptor 4	GRM4	Q14833
Metabotropic glutamate receptor 5	GRM5	P41594
Metabotropic glutamate receptor 6	GRM6	O15303
Metabotropic glutamate receptor 7	GRM7	Q14831
Metabotropic glutamate receptor 8	GRM8	O00222
Taste receptor 1 member 1	TAS1R1	Q7RTX1
Taste receptor 1 member 2	TAS1R2	Q8TE23
Taste receptor 1 member 3	TAS1R3	Q7RTX0

[0065] Described herein are GPCR binding domains, wherein the GPCR binding domains are designed based on surface interactions between a GPCR ligand and the GPCR. In some instances, the ligand is a subatomic particle (*e.g.*, a photon), an ion, an organic molecule, a peptide, and a protein. Non-limiting examples of ligands which can be bound by a GPCR include (-)-adrenaline, (-)-noradrenaline, (lyso)phospholipid mediators, [des-Arg10]kallidin, [des-Arg9]bradykinin, [des-Gln14]ghrelin, [Hyp3]bradykinin, [Leu]enkephalin, [Met]enkephalin, 12-hydroxyheptadecatrienoic acid, 12R-HETE, 12S-HETE, 12S-HPETE, 15S-HETE, 17 β -estradiol, 20-hydroxy-LTB₄, 2-arachidonoylglycerol, 2-oleoyl-LPA, 3-hydroxyoctanoic acid, 5-hydroxytryptamine, 5-oxo-15-HETE, 5-oxo-ETE, 5-oxo-ETrE, 5-oxo-ODE, 5S-HETE, 5S-HPETE, 7 α ,25-dihydroxycholesterol, acetylcholine, ACTH, adenosine

diphosphate, adenosine, adrenomedullin 2/intermedin, adrenomedullin, amylin, anandamide, angiotensin II, angiotensin III, annexin I, apelin receptor early endogenous ligand, apelin-13, apelin-17, apelin-36, aspirin triggered lipoxin A4, aspirin-triggered resolvin D1, ATP, beta-defensin 4A, big dynorphin, bovine adrenal medulla peptide 8-22, bradykinin, C3a, C5a, Ca²⁺, calcitonin gene related peptide, calcitonin, cathepsin G, CCK-33, CCK-4, CCK-8, CCL1, CCL11, CCL13, CCL14, CCL15, CCL16, CCL17, CCL19, CCL2, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, CCL3, CCL4, CCL5, CCL7, CCL8, chemerin, chenodeoxycholic acid, cholic acid, corticotrophin-releasing hormone, CST-17, CX3CL1, CXCL1, CXCL10, CXCL11, CXCL12 α , CXCL12 β , CXCL13, CXCL16, CXCL2, CXCL3, CXCL5, CXCL6, CXCL7, CXCL8, CXCL9, cysteinyl-leukotrienes (CysLTs), uracil nucleotides, deoxycholic acid, dihydrosphingosine-1-phosphate, dioleoylphosphatidic acid, dopamine, dynorphin A, dynorphin A-(1-13), dynorphin A-(1-8), dynorphin B, endomorphin-1, endothelin-1, endothelin-2, endothelin-3, F2L, Free fatty acids, FSH, GABA, galanin, galanin-like peptide, gastric inhibitory polypeptide, gastrin-17, gastrin-releasing peptide, ghrelin, GHRH, glucagon, glucagon-like peptide 1-(7-36) amide, glucagon-like peptide 1-(7-37), glucagon-like peptide 2, glucagon-like peptide 2-(3-33), GnRH I, GnRH II, GRP-(18-27), hCG, histamine, humanin, INSL3, INSL5, kallidin, kisspeptin-10, kisspeptin-13, kisspeptin-14, kisspeptin-54, kynurenic acid, large neuromedin N, large neurotensin, L-glutamic acid, LH, lithocholic acid, L-lactic acid, long chain carboxylic acids, LPA, LTB₄, LTC₄, LTD₄, LTE₄, LXA₄, Lys-[Hyp³]-bradykinin, lysophosphatidylinositol, lysophosphatidylserine, Medium-chain-length fatty acids, melanin-concentrating hormone, melatonin, methylcarbaryl PAF, Mg²⁺, motilin, N-arachidonoylglycine, neurokinin A, neurokinin B, neuromedin B, neuromedin N, neuromedin S-33, neuromedin U-25, neuronostatin, neuropeptide AF, neuropeptide B-23, neuropeptide B-29, neuropeptide FF, neuropeptide S, neuropeptide SF, neuropeptide W-23, neuropeptide W-30, neuropeptide Y, neuropeptide Y-(3-36), neurotensin, nociceptin/orphanin FQ, N-oleoylethanolamide, obestatin, octopamine, orexin-A, orexin-B, Oxysterols, oxytocin, PACAP-27, PACAP-38, PAF, pancreatic polypeptide, peptide YY, PGD₂, PGE₂, PGF₂ α , PGI₂, PGJ₂, PHM, phosphatidylserine, PHV, prokineticin-1, prokineticin-2, prokineticin-2 β , prosaposin, PrRP-20, PrRP-31, PTH, PTHrP, PTHrP-(1-36), QRFP43, relaxin, relaxin-1, relaxin-3, resolvin D1, resolvin E1, RFRP-1, RFRP-3, R-spondins, secretin, serine proteases, sphingosine 1-phosphate, sphingosylphosphorylcholine, SRIF-14, SRIF-28, substance P, succinic acid, thrombin, thromboxane A₂, TIP39, T-kinin, TRH, TSH, tyramine, UDP-glucose, uridine diphosphate, urocortin 1, urocortin 2, urocortin 3, urotensin II-related peptide, urotensin-II, vasopressin, VIP, Wnt, Wnt-1, Wnt-10a, Wnt-10b, Wnt-11, Wnt-16, Wnt-2, Wnt-2b, Wnt-3,

Wnt-3a, Wnt-4, Wnt-5a, Wnt-5b, Wnt-6, Wnt-7a, Wnt-7b, Wnt-8a, Wnt-8b, Wnt-9a, Wnt-9b, XCL1, XCL2, Zn²⁺, α -CGRP, α -ketoglutaric acid, α -MSH, α -neoendorphin, β -alanine, β -CGRP, β -D-hydroxybutyric acid, β -endorphin, β -MSH, β -neoendorphin, β -phenylethylamine, and γ -MSH.

[0066] Sequences of GPCR binding domains based on surface interactions between a GPCR ligand and the GPCR are analyzed using various methods. For example, multispecies computational analysis is performed. In some instances, a structure analysis is performed. In some instances, a sequence analysis is performed. Sequence analysis can be performed using a database known in the art. Non-limiting examples of databases include, but are not limited to, NCBI BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi), UCSC Genome Browser (genome.ucsc.edu/), UniProt (www.uniprot.org/), and IUPHAR/BPS Guide to PHARMACOLOGY (guidetopharmacology.org/).

[0067] Described herein are GPCR binding domains designed based on sequence analysis among various organisms. For example, sequence analysis is performed to identify homologous sequences in different organisms. Exemplary organisms include, but are not limited to, mouse, rat, equine, sheep, cow, primate (*e.g.*, chimpanzee, baboon, gorilla, orangutan, monkey), dog, cat, pig, donkey, rabbit, fish, fly, and human.

[0068] Following identification of GPCR binding domains, libraries comprising nucleic acids encoding for the GPCR binding domains may be generated. In some instances, libraries of GPCR binding domains comprise sequences of GPCR binding domains designed based on conformational ligand interactions, peptide ligand interactions, small molecule ligand interactions, extracellular domains of GPCRs, or antibodies that target GPCRs. Libraries of GPCR binding domains may be translated to generate protein libraries. In some instances, libraries of GPCR binding domains are translated to generate peptide libraries, immunoglobulin libraries, derivatives thereof, or combinations thereof. In some instances, libraries of GPCR binding domains are translated to generate protein libraries that are further modified to generate peptidomimetic libraries. In some instances, libraries of GPCR binding domains are translated to generate protein libraries that are used to generate small molecules.

[0069] Methods described herein provide for synthesis of libraries of GPCR binding domains comprising nucleic acids each encoding for a predetermined variant of at least one predetermined reference nucleic acid sequence. In some cases, the predetermined reference sequence is a nucleic acid sequence encoding for a protein, and the variant library comprises sequences encoding for variation of at least a single codon such that a plurality of different variants of a single residue in the subsequent protein encoded by the synthesized nucleic acid are generated by

standard translation processes. In some instances, the libraries of GPCR binding domains comprise varied nucleic acids collectively encoding variations at multiple positions. In some instances, the variant library comprises sequences encoding for variation of at least a single codon in a GPCR binding domain. In some instances, the variant library comprises sequences encoding for variation of multiple codons in a GPCR binding domain. An exemplary number of codons for variation include, but are not limited to, at least or about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 225, 250, 275, 300, or more than 300 codons.

[0070] Methods described herein provide for synthesis of libraries comprising nucleic acids encoding for the GPCR binding domains, wherein the libraries comprise sequences encoding for variation of length of the GPCR binding domains. In some instances, the library comprises sequences encoding for variation of length of at least or about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 225, 250, 275, 300, or more than 300 codons less as compared to a predetermined reference sequence. In some instances, the library comprises sequences encoding for variation of length of at least or about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, or more than 300 codons more as compared to a predetermined reference sequence.

[0071] Following identification of GPCR binding domains, the GPCR binding domains may be placed in scaffolds as described herein. In some instances, the scaffolds are immunoglobulins. In some instances, the GPCR binding domains are placed in the CDR-H3 region. GPCR binding domains that may be placed in scaffolds can also be referred to as a motif. Scaffolds comprising GPCR binding domains may be designed based on binding, specificity, stability, expression, folding, or downstream activity. In some instances, the scaffolds comprising GPCR binding domains enable contact with the GPCRs. In some instances, the scaffolds comprising GPCR binding domains enables high affinity binding with the GPCRs. Exemplary amino acid sequences of GPCR binding domains are described in **Table 2**.

Table 2. GPCR amino acid sequences

SEQ ID NO	GPCR	Amino Acid Sequence
1	CXCR4	MEGISIYTS DNYTEEMSGDYDSMKEPCFREENANFNKIFLPTIYSIIF LTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLFVITLPFWA VDAVANWYFGNFLCKAVHVIYTVNLYSSVLILAFISLDRYLAIVHA TNSQRPRKLLAEKVVYVGVWIPALLLTIPDFIFANVSEADDRYICDR FYPNDLWVVVFQFQHIMVGLILPGIVILSCYCIISKLSHSGHQRK

		ALKTTVILILAFFACWLPYYIGISIDSFILLEIKQGCEFENTVHKWISIT EALAFFHCCLNPILYAFLGAKFKTSAQHALTSVSRGSSLKILSKGKR GGHSSVSTESESSSFHSS
2	CCR4	MNPTDIADTTLDESIYSNYYLYESIPKPCTKEGIKAFGELFLPPLYSL VFVFGLLGNSVVVLVLFKYKRLRSMTDVYLLNLAISDLLFVFSLPF WGYAADQWVFGGLCKMISWMYLVGFYSGIFFVMLMSIDRYLAI VHAVFSRLARTLTYGVITSLATWSVAVFASLPGFLFSTCYTERNHTY CKTKYSLNSTTWKVLSSLEINILGLVIPLGIMLFCYSMIIRTLQHCKN EKKNKAVKMIFAVVVLFLGFWTPYNIVLFLETLVELEVLQDCTFER YLDYAIQATETLAFVHCCLNPIIYFFLGEKFRKYILQLFKTCRGLFVL CQYCGLLQIYSADTPSSSYTQSTMDHDLHDAL
3	GCGR	MPPCQPQRPLLLLLLLACQPQVPSAQVMDLFEKWKLYGDQCHH NLSLLPPPTLVCNRTFDKYSCWPDTPANTTANISCPWYLPWHHKV QHRFVFKRCGPDGQWVRGPRGQPWDASQCQMDGEEIEVQKEVA KMYSSFQVMYTVGYSLGALLLALAILGGLSKLHCTRNAIHANLF ASFVLKASSVLVIDGLLRTRYSKIGDDLSTWLDGAVAGCRVA AVFMQYGIVANYCWLLVEGLYLHNLLGLATLPERSFFSLYLIGIGWG APMLFVVPWAVVKCLFENVQCWTSNDNMGFWWILRFPVFLAILIN FFIFVRIVQLLVAKLRARQMHTDYKFRLLAKSTLTLIPLLGVEHVVF AFVTDEHAQGTLSAKLFFDLFLSSFQGLLVAVLYCFLNKEVQSEL RRRWHRWRLGKVLWEERNTSNHRASSSPGHGPPSKELQFGRGGGS QDSSAETPLAGGLPRLAESPF
4	mGluR5	MVLLILSVLLLKEDVRGSAQSSERRVVAHMPGDIIGALFSVHHQP TVDKVHERKCGAVREYQGIQRVEAMLHTLERINSDPTLLPNITLGC EIRDSCWHSVALEQSIIEFIRDSLISSEEEGLVRCVDGSSSSFRSKKP IVGVIGPGSSVAIQVQNLQLFNIPQIAYSATSMDSLTKLTKYFMR VVPDAQQARAMVDIVKRYNWTYVSAVHTEGNYGESGMEAFKD MSAKEGICIAHSYKIYSNAGEQSFDKLLKLTSHLPKARVVACFCEG MTVRGLLMAMRRLGLAGEFLLGSDGWADRYDVTGQREAVG GITIKLQSPDVKWFDYYLKLRPETNHRNPWFQEFWQHRFQCRLEG FPQENSKYNTCNSSLTKTHHVQDSKMGFVINAIYSMAYGLHNM QMSLCPGYAGLCDAMKPIDGRKLLESLMKTNFTGVSGDTILFDENG DSPGRYEIMNFKEMGKDYFDYINVGSWDNGELKMDDDEVWSKKS NIIRSVCEPCEKGQIKVIRKGEVSCCWTCTPCKENYVFDEYTCKA CQLGSWPTDDLGTGCDLIPVQYLRWGDPEPIAAVVFACLGLLATLFV TVVFIIYRDTVPVKSSSRELCYIILAGICLGYLCTFCLIAKPKQIYCYL QRIGIGLSPAMSYSALVTKTNRIARILAGSKKKICTKKPRFMSACAQ LVIAFILICIQLGIIVALFIMEPPDIMHDYPSIREVYLICNTTNLGVVTP LGYNGLLILSCTFYAFKTRNVANFNEAKYIAFTMYTTTCIWLAFVPI YFGSNYKIITMCFVSLSATVALGCMFVPKVYIILAKPERNVRSFTT STVVRMHVGDGKSSSAASRSSLVNLWKRGGSSGETLRYKDRRLA QHKSEIECFTPKGSMGNGGRATMSSSNGKSVTWAQNEKSSRGQHL WQRLSIHINKKENPNQTAVIKPFKSTESRGLGAGAGAGGSAGGVG ATGGAGCAGAGPGGPESPDAGPKALYDVAEAEHFAPARPRSPSP ISTLSHRAGSASRTDDDVPSLHSEPVARSSSSQGSLEQISSVTRFT ANISELNSMMLSTAAPSPGVGAPLCSSYLIPKEIQLPTTMTTFAEIQP LPAIEVTGGAQPAAGAQAAGDAARESAPAGPEAAAAPDLEELVA LTPSPFRDSVDSGSTTPNSPVSEALCIPSSPKYDTLIIRDYTSSSSL
5	GLP-1R	RPQGATVSLWETVQKWREYRRQCQRSLTEDPPPATDLFCNRTFDE YACWPDGEPGSFVNVSCPWYLPWASSVPQGHVYRFCTAEGWLWQ KDNSSLPWRDLSECEESKRGERSSPEEQLLFLYIIYTVGYALSFSALV

		IASAILLGFRHLHCTRNYIHLNLFASFILRALSVFIKDAALKWMYSTA AQQHQWDGLLSYQDSLSCRLVFLLMQYCVAANYWLLVEGVYLY TLLAFSVLSEQWIFRLYVSIGWGVPLLFVVPWGVKLYEYDEGCWT RNSNMNYWLIIRLPILFAIGVNFLIFVRVICIVVSKLKANLMCKTDIK CRLAKSTLTLIPLLGTHEVIFAFVMDEHARGTLRFIKLFTELSFTSFQ GLMVAILYCFVNNEVQLEFRKSWERWRLEHLHIQRDSSMKPLKCPT SSLSSGATAGSSMYTATCQASCS
6	GABAB	MLLLLLLAPLFLRPPGAGGAQTPNATSEGCQIIHPPWEGGIRYRGLT RDQVKAINFLPVDYIEIYVCRGEREVVGPKVRKCLANGSWTDMDT PSRCVRICSKSYLTLENGKVFLTGGDLPALD GARVDFRCDPDFHLV GSSRSICSQQWSTPKPHCQVNRTPHSERRAVYIGALFPMSSGGWPG GQACQPAVEMALEDVNSRRDILPDYELKLIHHSKCDPGQATKYL YELLYNDPIKILMPGCSSVSTLVAEAARMWNLIVLSYSSSPALS RQRFTFFRTHPSATLHNPTRVKLFKKGWKKIATIQTTVEFTSTL DDLEERVKEAGIEITFRQSFFSDPAVPVKNLKRQDARIIVGLFYETEA RKVFCEVYKERLFGKKYVWFLIGWYADNWFKIYDPSINCTVDEMT EAVEGHITTEIVMLNPANTRISNMTSQEFVEKLT KRLKRHPREETGG FQEAPLAYDAIWALALALNKTSGGGGRSGVRLEDFNYNNQTITDQI YRAMNSSFEGVSGHVVDASGSRMAWTLIEQLQGGSYKKIGYYD STKDDLWSKTDKWIGGSPPADQTLVIKTRFLSQKLFISVSVLSSLG IVLAVVCLSFNIYNHVRVYIQNSQPNLNNLTAVGCSLALAAVFPLGL DGYHIGRNQFPFVCQARLWLLGLGFSLGYSMTKIWWVHTVFTK KEEKKEWRKTLEPWKLYATVGLLVGMVDVLTALWQIVDPLHRTIE TFAKEPKEDIDVSILPQLEHCSSRKMNTWLGIFYGYKGLLLLLGIFL AYETKSVSTEKINDHRAVGMAIYNVAVLCLITAPVTMILSSQQDAA FAFASLAIVFSSYITLVLFVPMRRLITRGEWQSEAQDTMKTGSST NNNEEEKSRLLKENRELEKIIAEKEERVSELRHQLQSRQQLRSRRH PPTPEPSGGLPRGPPEPPDRLSCDGSRVHLLYK
7	OPRM1	MDSSAAPTNASNCTDALAYSSCSPAPSPGSWVNLSHLDGNLSDPCG PNRTDLGGRDSLCPPTGSPSMITAITIMALYSIVCVVGLFGNFLVMY VIVRYTKMKTATNIYIFNLALADALATSTLPFQSVNYLMGTWPFGTI LCKIVISIDYNYMFTSIFTLCTMSVDRIYAVCHPVKALDFRTPRNAKI INVCNWILSSAIGLPVMFMATTKYRQGSIDCTLTFSHPTWYWENLL KICVFIFAFIMPVLIITVCYGLMILRLKSVRMLSGSKEKDRNLRRITR MVLVVVAVFIVCWTPIHIVVIKALVTIPETTFQTVSWHFCIALGYTN SCLNPVLYAFLDENFKRCFRECIPTSSNIEQQNSTRIRQNTRDHPST ANTVDRTNHQLENLEAETAPLP
8	OPRK1	MDSPIQIFRGEPTCAPSACLPPNSSAWFPGWAEPDSNGSAGSEDA QLEPAHISPAIPVIITAVYSVVFVGLVGNLSVMFVIIRYTKMKTATN IYIFNLALADALVTTTTPFQSTVYLMNSWPFGDVLCKIVISIDYNYM FTSIFTLTMSVDRIYAVCHPVKALDFRTPKAKIINICIWLLSSSVGI SAIVLGGTKVREDVDVIECSLQFPDDDYSWWDLFMKICVFIFAFVIP VLIIIVCYTLMILRLKSVRLSGSREKDRNLRRITRLVVLVVAVFVVC WTPIHIFILVEALGSTSHSTAALSSYYFCIALGYTNSSLNPILYAFLE NFKRCFRDFCFPLKMRMERQSTSRRVNTVQDPAYLRDIDGMNKPV
9	C5aR	MDSFNYYTPDYGHYDDKDTLDLNTVPDKTSNTLRVPDILALVIFAV VFLVGVGLGNALVWVTAFAEAKRTINAIWFLNLAVADFLSCLALPIL FTSIVQHHPFPGGAACSLPSLILLNMYASILLLATISADRFLLVFKP IWCQNFRGAGLAWIACAVAWGLALLLTIPSFLYRVVREEYFPPKVL CGVDYSHDKRRERAVAIVRLVLGFLWPLLTLICYTFILLRTWSRRA TRSTKTLKVVVAVVASFFIFWLPYQVTGIMMSFLEPSSPTFLLKKL

		DSLCSFAYINCCINPIIYVAGQGFGRLRKSLSLLRNVLTEESVV RESKSFTRSTVDMAQKTQAV
10	CGRP	ELEESPEDSIQLGVTRNKIMTAQYECYQKIMQDPIQQAEGVYCNRT WDGWLCWNDVAAGTESMQLCPDYFQDFDPSEKVTKICDQDGNWF RHPASNRTWTNNTQC NVNTHEKVK TALNLFYLTIIHGHSIASLLIS LGIFFYFKSLSCQRITLHKNLFFSFVCNSVVTIIHLTAVANNQALVAT NPVSCKVSQFIHLYLMGCNYFWMLCEGIYLHTLIVVAVFAEKQHL MWWWFLGWGFPLIPACIHAIARSLYYNDNCWISSDTHLLYIIHGPIG AALLVNLFFLLNIVRVLITKLKVTHQAESNLYMKAVRATLILVPLLG IEFVLIPWRPEGKIAEEVYDYIMHILMHFQGLLVSTIFCFFNGEVQAI LRRNWNQYKIQFGNSFSNSEALRSASYTVSTISDGPYSHDCPSEHL NGKSIHDIENVLLKPENLYN
11	M1 muscarinic	MNTSAPPAVSPNITVLA PGKGPWQVAFIGITTGLLSLATVTGNLLVL ISFKVNTELKTVNNYFLLSLACADLIIGTFSMNLYTTYLLMGHWAL GTLACDLWLALDYVASNASVMNLLISFDYFSVTRPLSYRAKRT RRAALMIGLAWLVSVLWAPAILFWQYLVGERTVLAGQCYIQFLS QPIITFGTAMAAFYLPVTVMCTLYWRIYRETERARELAAALQGSETP GKGGGSSSSSERSQPGAEGSPETPPGRCCRCCRAPRLQAYSWKEEE EEDEGSME SLTSSEGEPEGSEVVIKMPMVDPEAQAPTQPPRSPNT VKRPTKKGRDRAGKGQKPRGKEQLAKRKTFSLVKEKKAARTLSAI LLAFILTWTYPYNIMVLVSTFCKDCVPETLWELGYWLCYVNSTINPM CYALCNKA FRDTRLLLLCRWDKRRWRKIPKRP GSVHRTPSRQC
12	M4 muscarinic	MANFTPVNGSSGNQSVRLVTSSSHNRYETVEMVFIATVTGSLSLVT VVGNI LVMLSIKVNRQLQTVNNYFLFSLACADLIIGAFSMNLYTVYI IKGYWPLGAVVCDLWLALDYVVSNASVMNLLISFDYFCVTKPLT YPARRTTK MAGLMIAAAWVLSFVLWAPAILFWQFVVGKRTVPDN QCFIQFLSNPAVTFGTAAAFYLPVVIMTVLYIHISLASRSRVHKHRP EGPKEKKAKTALFLKSPLMKQSVKKPPPGEAA REELRNGKLEEAPP PALPPPPRPVADKDTSNESSSGSATQNTKERPA TELSTTEATTPAMP APPLQPRALNPASRWSKIQIVTKQTGNECVTAIEIVPATPAGMRPAA NVARKFASIARNQVRKKRQMAARERKVTRTIFAILLAFILTWTYPYN VMVLVNTFCQSCIPDTVWSIGYWLCYVNSTINPACYALCNATFKKT FRHLLLCQYRNIGTAR
13	CCR2	MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQLLPP LYSLVFIFGFGVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLP LWAHSAANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIV HAVFALKARTVTFGVVTSVITWLAVFASVPGIIFTKCQKEDSVYV CGPYFPRGWNNFHTIMRNILGLVPLLMVICYSGLKTLRLCRNEK KRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSNCESTS QLD QATQVTETLGMTHCCINPIIYAFVGEKFRSLFHIALGCRIAPLQKPVC GGPGVRPGKNVKVTTQGLLDGRGKGK SIGRAPEASLQDKEGA
14	CCR9	MTPTDFTSPIPNMADDYGSESTSSMEDYVNFNFTDFYCEKNNVRQF ASHFLPLYWL VFIVGALGNSLVILVYWYCTR VKTMTDMFLLNLAI ADLLFLVTL PFWAIAAADQWK FQTFMCKVVNSMYKMNFYSCVLLI MCISVDRYIAIAQAMRAHTWREKRLLYSKMVCFTIWLAAALCIPE ILYSQIKEESGIAICTMVYPSDESTKLKSAVLT LKVLGFFLPFV VMA CCYTIIHTLIQAKKSSKH KALKVTITVLT VLVLSQFPYNCILLVQTID AYAMFISNCAVSTNIDICFQVTQTIAFFHSCLNPVLYV FVGERFRRDL VKTLKNLGCISQAQWVSFTRREGSLKLSSMLLETTSGALS L
15	GPR174	MPANYTCTRPDGDNTDFRYFIYAVTYTVILVPGLIGNILALWV FYG YMKETKRAVIFMINLA IADLLQVLSLPLRIFYL NHDWPFPGPLCMF

		CFYLKYVNMYASIFYFLVCISVRRFWFLMYPFRFHDCKQKYDLYISIA GWLIIICLACVLFPLLRTSDDTSGNRTKCFVDLPTRNVNLAQSVVMM TIGELIGFVTPLLIVLYCTWKTIVLSLQDKYPMAQDLGEKQKALKMI LTCAGVFLICFAPYHFSFPLDFLVKSNEIKSCLARRVILIFHSVALCLA SLNSCLDPVIYYFSTNEFRRLSRQDLHDSIQLHAKSFVSNHTASTM TPELC
16	MASP-2	TPLGPKWPEPVFGRLASPGFPGHEYANDQERRWTLTAPPGYRLRLYF THFDLELSHLCEYDFVKLSSGAKVLATLCGQESTDTERAPGKDTFY SLGSSLDITFRSDYSNEKPFTGFEEFYAAEDIDECQVAPGEAPTCDH HCHNHLGGFYCSCRAGYVLHRNKRTCSALCSGQVFTQSRGELSSPE YPRYPKLSSCTYSISLEEGFSVILDFVESFDVETHPETLCPYDFLKIQ TDREEHGPFCGKTLPHRIETKSNTVTITFVTDESGDHTGWKIHYTST AQPCPYPMAPPNGHVSPVQAKYILKDSFSIFCETGYELLQGHPLKS FTAVCQKDGSDWRPMPACSIIVDCGPPDDLPSGRVEYITGPVCTTYK AVIQYSCEETFYTMKVNDGKYVCEADGFWTSSKGEKSLPVCEPVC GLSARTTGGRYGGQKAKPGDFPWQVLILGGTTAAGALLYDNWVL TAAHAVYEQKHDA SALDIRMGTCLKRLSPHYTQAWSEAVFIHEGYT HDAGFDNDIALIKLNNKVINSNITPICLPRKEAESFMRTDDIGTASG WGLTQRGFLARNLMYVDIPIVDHQCTAAEYKPPYPRGSVTANML CAGLES GGKDS CRGDSGGALVFLDSETERWFGGIVSWGSMNCGE AGQYGVYTKVINYIPWIENIISDF
17	CCR5	MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGN MLVILILINCKRLKSMTDIYLLNLAISDLFLLTVPFWAHYAAAQWD FGNTMCQLLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVT FGVVTSVITWVAVFASLPGIIFTRSQKEGLHYTCSSHPYSQYQFW KNFQTLKIVILGLVPLLMVICYSILKTLRLCRNEKKRHRVRLIF TIMIVYFLFWAPYNIVLLLNTFQEFFGLNCCSSSNRLDQAMQVTETL GMTHCCINPIYAFVGEKFRNYLLVFFQKHIKRCKCCSIFQQEAPPE RASSVYTRSTGEQEISVGL
18	FSHR	CHHRICHCSNRVFLCQESKVTEIPSDLPRNAIELRFVLTCLRVIQKGA FSGFGDLEKIEISQNDVLEIEADVFSNPKLHEIRIEKANNLLYINPE AFQNLPNLQYLLISNTGIKHLPDVHKIHSKQVLLDIQDNINIHTIER NSFVGLSFESVILWLNKNGIQEIHNCANFTQLDELNLSNNEEL PNDVFHSGASGPVILDISRTIHSLSYGLNLLKLRARSTYNLKKLPT LEKLVALMEASLTYPHCCAFANWRRQISELHPICNKSILRQEVYD MTQARGQRSSLAEDNESSYSRGFDMTYTEFDYDLCNEVVDTVCSP KPD AFN PCEDIMGYNLRVLIWFISILAITGNIIVLVILTTSQYKLTVP FLMCNLAFADLCIGIYLLIASVDIHTKSQYHNYAIDWQTGAGCDA AGFFT VFA SELSVYTLTITLERWHTITHAMQLDCKVQLRHAASVM VMGWIFAFAAALFPIFGISSYMKVSICLPMDIDSPLSQLYVMSLLVL NVLA FVVICGCIYIHYLTVRNPNI VSSSDTRI AKRMAMLI FTDFLCM APISFFAISASLVPLITVSKAKILLVLFHPINSCANPFLYAIFTKNFR DFFILLSKCGCYEMQAQIYRTETSSTVHNTHPRNGHCSSAPRV TNGS TYILVPLSHLAQN
19	mGluR2 PAM	EGPAKKVLTLEGDLVLGGFLFPVHQKGGPAEDCGPVNEHRGIQRLEA MLFALDRINRDPHLLPGVRLGAHILDSCSKDTHALEQALDFVRASLS RGADGSRHICPDGSYATHGDAPTAITGVIGGSYSDVSIQVANLLRLF QIPQISYASTSAKLSKSDRYDYFARTVPPDFQAKAMAEILRFFNWT YVSTVASEGDYGETGIEAFELEARARNICVATSEKVGGRAMSRAAFE GVVRALLQKPSARVAVLFTREDAARELLAASQRLNASFTWVASDG WGALESVVAGSEGAAEGAITIELASYPISDFASYFQSLDPWNNSRNP

		WFREFWEQRFRCFSFRQRDCAAHSLRAVPFEQESKIMFVVNAVYAM AHALHNMHRALCPNTTRLCDAMRPVNGRRLYKDFVLNVKFDAPF RPADTHNEVRFDREFDGGIGRYNIFTYLRAGSGRYRYQKVGYWAEG LTLDTSLIPWASPSAGPLPASRCSEPCLQNEVKSVPGEVCCWLCIP CQPYEYRLDEFTCADCGLYWPNASLTGCFELPQEYIRWGDAAV GPVTIACLGALATLFLVGVFVRHNATPVVKASGRELCYILLGGVFL CYCMTFIFIAPKSTAVCTLRRLGLGTAFSVCYSALLTKTNRIARIFGG AREGAQRPRFISPASQVAICLALISGQLLIVVAWLVEAPGTGKETA PERREVVTLCNHRDASMLGSLAYNVLLIALCTLYAFKTRKCPENF NEAKFIGFTMYTTCIWLAFLPIFYVTSSDYRVQTTTMCVSVSLSGSV VLGCLFAPKLHIILFQPQKNVSHRAPTSRFGSAAARASSSLGQSGS SQFVPTVCNGREVVDSTTSSL
20	mGluR3	LGDHNFLRREIKIEGDLVLGGFLPINEKGTGTEECGRINEDRGIQRL AMLFAIDEINKDDYLLPGVKLGVLHLDTCSDTYALEQSLEFVRASL TKVDEAEYMCPDGSYAIQENIPLLIAGVIGGSYSSVSQVANLLRFLQ IPQISYASTSAKLSDKSRYDYFARTVPPDFYQAKAMAEILRFFNWTY VSTVASEGDYGETGIEAFEQEARLRNICIATAEKVGRSNIRKSYDSVI RELLQKPNARVVVLFMRSDDSRELIAAASRANASFTWVASDGGWA QESIIGSEHVAYGAITLELASQPVRQFDYFQSLNPNYNNHRNPWFR DFWEQKFQCSLQNKRNHRRVCDKHLAIDSSNYEQESKIMFVVNAV YAMAHALHKMQRTLCPNTTKLCDAMKILDGKKLYKDYLLKINFTA PFNPNKDADSIVKFDTFGDGMGRYNVFNFNQNVGGKYSYLKVGHW AETLSLDVNSIHWSRNSVPTSQCSDPCAPNEMKMNMQPGDVCCWICI PCEPYEYLADEFTCMDCGSGQWPTADLTGCYDLPEDYIRWEDAWA IGPVTIACLGFMCTCMVVTVFIKHNTPLVKASGRELCYILLFGVGL SYCMTFFFIAPKSPVICALRRLGLGSSFAICYSALLTKTNCIARIFDGV KNGAQRPKFISPSSQVFICLGLILVQIVMVSVWLILEAPGTRRYTLAE KRETVILKCNVKDSSMLISLTYDVILVILCTVYAFKTRKCPENFNEA KFIGFTMYTTCIWLAFLPIFYVTSSDYRVQTTTMCISVSLSGFVVLG CLFAPKVHIILFQPQKNVTHRLHLNRFSVSGTGTTYSQSSASTYVP TVCNGREVLDSTTSSL
21	mGluR4	KPKGHPHMNSIRIDGDTLGGFLFPVHGRGSEGKPCGELKKEKGIHRL EAMLFALDRINNDPDLNITLGARILDTCSDRTHALEQSLTFVQALI EKDGTVEVRCGSGGPPITKPERVVGVIASGSSVSIMVANILRLFKIP QISYASTAPDLSDNSRYDFFSRVVPSTYQAQAMVDIVRAKWNV VSTVASEGSYGESGVEAFIQKSREDGGVCIAQSVKIPREPKAGEFDKI IRRLLETSNARAVIIFANEDDIRRVLEAARRANQTGHFFWMGSDSW GSKIAPVLHLEEVAECAVTILPKRMSVRGFDYFSSRTLNNRRNIW FAEFWEDNFHCKLSRHALKKGSHVKKCTNRERIGQDSAYEQEGKV QFVIDAVYAMGHALHAMHRDLCPGRVGLCPRMDPVDGTQLLKYI RNVNFSGIAGNPVTFNENGDAAGRYDIYQYQLRNDSEYKVGISWT DHLHLRIERMHWPGSGQQLPRISCSLPCQPGERKKTVKGMPCWH CEPCTGYQYQVDYRTCKTCPCYDMRPTENRTGCRPIPIKLEWGPW AVLPLFLAVVGIAATLFVVITFVRYNDTPIVKASGRELSYVLLAGIFL CYATTFLMIAEPDLGTCSLRRIFLGLGMSISYAALLTKTNRIYRIFEQ GKRSVSAPRFISPASQLAITFSLISLQLLGICVWFVVDPSHSVVDQD QRTLDPRFARGVLKCDISDLSLICLLGYSMLLMVTCTVYAIKTRGVP ETFNEAKPIGFTMYTTCIVWLAFIPIFFGTSQSADKLYIQTTTLTVSVS LSASVSLGMLYMPKVYIILFHPEQNVPKRKRSLKAVVTAATMSNKF TQKGNFRPNGEAKSELNLEAPALATKQTYVTYTNHAI
22	mGluR7	QEMYAPHSIRIEGDVTLGGFLFPVHAKGPSGVPCGDIKRENGIHRLEA

		MLYALDQINSDPNLLPNVTLGARILDTC SRDTYALEQSLTFVQALIQ KDTSDVRCNTNGEPPVFKPEKVVG VIGASGSSVSIMVANILRLFQIP QISYASTAPELSDDRRYDFFSRVVPD SFQAQAMVDIVKALGWNYV STLASEGSYGEKGVESFTQISKEAGGLCIAQSVRIPQERKDRTIDFDR IHKQLDTPNSRAVVIFANDEDIKQILAAAKRADQVGHFLWVGSDS WGSKINPLHQHEDIAEGAITIQPKRATVEGFDAYFTSR TLENNRRNV WFAEYWEENFNCKLTISGSKKEDTDRKCTGQERIGKDSNYEQEGK VQFVIDAVYAMAHALHHMNKDLCADYRGVCP EMEQAGGKKLLK YIRNVNFNGSAGTPVMFNKNGDAPGRYDIFQYQTTNTSNPGYRLIG QWTDELQNLIEDMQWGKGVREIPASVCTLPCKPGQRKKTKQKTPC CWTCEPCDGYQYQFDEMTCQHCPYDQRPENRTGCQDIP IIKLEWH SPWAVIPVFLAMLGIIATIFVMATFIRYNDTPIVRASGREL SYVLLTGI FLCYIITFLMIAKPDVAVCSFRRVFLGLGMCISYAALLTKTNRIYRIF EQGKKSVTAPRLISPTSQLAITSSLISVQLLG VFIWFGVDPPNIIDYDE HKTMNPEQARGVLKCDITDLQIICSLGY SILLMVTCTVYAIKTRGVP ENFNEAKPIGFTMYTTCIVWLAFIPIFFGTAQSAEKLYIQTTTLTISMN LSASVALGMLYMPKVYIIIFHPELNVQKRKRSFKAVVTAATMSSRL SHKPSDRPNGEAKTELCEVDPNSPAKKKYVSYNLVI
23	CXCR3	MVLEVSDHQVLNDAEVAALLENFSSSYDYGENESDSCCTSPPCPD FSLNFDRAFLPALYSLLFLLGLLGNGAVA AVLLSRRTALSSTDTFLL HLAVADTLLVLTPLWAVDAAVQWVFGSGLCKVAGALFNINFYA GALLLACISFDRLNIVHATQLYRRGPPARVLTCLAVWGLCLLFA LPDFIFLSAHHDERLNATHCQYNFPQVGR TALRVLQLVAGFLLPLL VMAYCYAHILAVLLVSRGQRRRLRAMRLVVVVV VAFALCWTPYHL VVLVDILMDLGALARNCGRESRVDVAKSVTSG LGYMHCCLNPLLY AFVGVKFRERMWMLLLRLGCPNQRGLQRQPSSSRDSSWSETSEA SYSGL
24	CCR8	MDYTLDLSVTTVTDYYPDIFSSPCDAELIQ TNGKLLAVFYCLLFV FSLLGNSLVILVLVVCCKLRSITDVYLLNLALSDLLVFVSFPQTYYL LDQWVFGTVMCKVVSGFYIYGFYSSMFFITLMSVD RYLAVVHAVY ALKVRTIRMGTTLCLAVWLTAIMATIPLL VFYQVASEDGVLQCYSF YNQQTWKWKIFTNFKMNLGLLIPFTIFMFCYIKILHQLKRCQNHNK TKAIRLV LIVVIASLLFWVPFNVVLF LTSLHSMHILDGCSISQQLTYA THVTEHSFTHCCVNPVIYAFVGEKFKKHLSEIFQKSCSQIFNYLGRQ MPRESCEKSSSCQQHSSRSSVDYIL
25	Adenosine A2a	MPIMGSSVYITVELAIAVLAILGNVLCWAVWLNSNLQNV TNYFV VSLAAADIAVGVLAIPIFAITISTGFCAACHGCLFIACFVLVLTQSSIFS LLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFAIGLTPMLGW NNGGQPKKEGKNHSQGCGEQVACL FEDVVP MNMVFNFACVL VPLLLMLGVYLRIFLAARRQLKQMESQPLPGERARSTLQKEVHA AK SLAIIVGLFALCWLP LHIINCFTFFCPDCSHAPLWLMYLAIVLSHTNS VYNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAAGTSARVLA AHG SDGEQVSLRLNGHPPGVWANGSAPHPERRPNGYALGLVSGGSAQE SQGNTGLPDVELLSHELKGVCEPPGLDDPLAQDGAGVS
26	Orexin OX1	MEPSATPGAQMGVPPGSRPSPVPPDYED EFLRYLWRDYLYPKQYE WVLIAAYVAVFVVALVGNTLVCLAVWRNHMRTVTNYFIVNL SL ADVLVTAICLPASLLVDITESWLF GHALCKVIPYLQAVSVSAVLT L SFIALDRWYAICHPLLFKSTARRARGSILGIWAVSLAIMVPQA AVM CSSVLPELANRTRLFSVCDERWADDLYPKIYHSCFFIVTYLAPLGLM AMAYFQIFRKLWGRQIPGTT SALVRNWKRPDQLGDLEQGLSGEPQ PRARAFLAEVKQMRARRKTAKMLMVLLVFALCYLPISVLNV LKR

		VFGMFRQASDREAVYACFTFSHWLVYANSAANPIIYNFLSGKFREQ FKAASFSCCLPGLGPCGSLKAPSPRSSASHKSLSLQSRCSISKISEHVVL TSVTTVLP
27	Orexin OX2	MSGTKLEDSPPCRNWSSASELNETQEPFLNPTDYDDEEFLRYLWRE YLHPKEYEWVLIAGYIIVFVVALIGNVLVCVAVWKNHHMRTVTNY FIVNLSLADVLVTITCLPATLVVDITETWFFGQSLCKVIPYLQTVSVS VSVLTLSIALDRWYAICHPLMFKSTAKRARNISIVIIWIVSCIIMIPQA IVMECSTVFPGLANKTTLFTVCDERWGGEIYPKMYHICFFLVTYMA PLCLMVLAYLQIFRKLWCRQIPGTSSVVQRKWKPLQPVSQPRGPGQ PTKSRMSAVAAEIKQIRARRKTARMLMIVLLVFAICYLPISILNVLKR VFGMFAHTEDRETVYAWFTFSHWLVYANSAANPIIYNFLSGKFREE FKAASFSCCLGVHHRQEDRLTRGRTSTESRKSLTTQISNFDNISKLSE QVVLTSISTLPAANGAGPLQNW
28	PAR-2	IQGTNRSSKGRSLIGKVDGTSHVTGKGVTVETVFSVDEFSASVLTGK LTTVFLPIVYTIVFVGLPSNGMALWVFLFRTKKKHPAVIYMANLA LADLLSVIWFPLKIAHYHIGNNWIYGEALCNVLIGFFYGNMYCSILF MTCLSVQRYWVIVNPMGHSRKKANIAIGISLAIWLLILLVTIPLYV KQTIFIPALNITTCHDVLPEQLLVGDMFNYFLSLAIGVFLFPAFLTAS AYVLMIRMLRSSAMDENSEKKRKRAIKLIVTVLAMYLICFTPSNLLL VVHYFLIKSQGQSHVYALYIVALCLSTLNSCIDPFVYFVSHDFRDH AKNALLCRSVRTVKQMQLVSLTSSKKHSRKSSSYSSSTTVKTSY
29	C3aR	MASFSAETNSTDLSQPWNEPPVILSMVILSLTFLGLPGNGLVLWV AGLKMQRVTNTIWFHLTLADLLCCLSLPFLAHLALQGQWPYGRF LCKLIPSIIVLNMFAVFLTAISLDRCLVVFKEPIWCQNHRNVGMAC SICGCIWVVAFVMCIPVFVYREIFTTDNHNRCGYKFGLSSSLDYPDF YGDPLENRSLENIVQPPGEMNDRLDPSSFQTNHDPWTVPTVFQQT FQRPSADSLPRGSARLTSQNLVSNVFKPADVVSPIKPSGFPEDHETS PLDNDAFLSTHLKLFPSASSNSFYSELPPQGFQDYNNLQFTDDDQ VPTPLVAITITRLVVGFLPSVIMIACYSFIVFRMQRGRFAKSTF RVAVVVVAVFLVCWTPYHIFGVLSLLTDPETPLGKTLMSWDHVCIA LASANSCFNPFYALLGKDFRKKARQSIQGILEAAFSEELTRSTHCP NNVISERNSTTV
30	LGR5	GSSPRSGVLLRGCPHCHCEPDGRMLLRVDCSDLGLSELPSNLSVFT SYLDLSMNNISQLLPNPLPSLRFLEELRLAGNALTYPKGAFGLYSL KVLMLQNNQLRHVPTEALQNLRLSLQSLRLDANHISYVPPSCFSGLH SLRHLWLDDNALTEIPVQAFRSLALQAMTLALNKIHHIPDYAFGN LSSLVVLHLHNNRIHSLGKKCFDGLHSLETLDLNNNLDEFPTAIRT LSNLKELGFHSNNIRSIPEKAFVGNPSLITIHFYDNPIQFVGRSAFQHL PELRTLTLNGASQITEFPDLTGANLESLLTGAQISSLPQTVCNQLP NLQVLDLSYNLLEDLPFSVCQKLQKIDLRHNEIYEIKVDTFQQLLS LRSLNLAWNKAIIHPNAFSTLPSLIKLDLSSNLLSFPITGLHGLTHL KLTGNHALQSLISSENFPELKVIEMPYAYQCCAFVCENAYKISNQ WNKGDNSSMDDLHKKDAGMFQAQDERDLEDLDFEEDLKALHS VQCSPSPGPFKPEHLLDGWLIRIGVWTIAVLALTCNALVTSTVFRS PLYISPIKLLIGVIAAVNMLTGVSSAVLAGVDAFTFGSFARHGAWW ENGVGCHVIGFLSIFASESSVFLTLAALERGFVSKYSAKFETKAPFS SLKVILLCALLALTMAAVPLLGGSKYGASPLCLPLPFGEPSTMGYM VALILLNSLCFLMMTIAYTKLYCNLDKGDLENIWDCSMVKHIALLL FTNCILNCPVAFLSFSSLINLTFISPEVIKFIILLVVVPLPACLNPLLYILF NPHFKEDLVSLRKQTYVWTRSKHPSLMSINSDDVEKQSCDSTQALV TFTSSSITYDLPPSSVPSPAYPVTESCHLSSVAFVPC

31	GPR101	MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVG NIVLALVLQRKPQLLQVTNRIFNLLVTDLLQISLVAPWVVSATSVPL FWPLNSHFCTALVSLTHLFAFASVNTIVVVSVDYRLSIIHPLSYPSKM TQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNALCSMIWGA SPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEV RVKDCVENEDEEGAEEKKEEFQDESEFRRQHEGEVKAKEGRMEAKD GSLKAKEGSTGTSESSVEARGSEEVRESSTVASDGSMEGKEGSTKV EENSMKADKGRTEVNQCSIDLGEDDMEFGEDDINFSEDDVEAVNIP ESLPPSRNSNSNPPLPRCYQCKAAKVIIFISYVLSLGPYCFLAVLA VWVDVETQVPQWVITIIWLFLLQCCIHYPYVGYMHKTIKKEIQDM LKKFFCKEKKPKEDSHPDLPGETEGGTEGKIVPSYDSATFP
32	GPR151	MLAAAFADSNSSSMNVSFALHFAAGGYLPSDSQDWRTHIIPALLVAV CLVGFVGNLCVIGILLHNAWKGKPSMIHSLILNLSLADLSLLFSAPI RATAYSKSVWDLGWVFCSSDWFIHTCMAAKSLTIVVVAKVCFM YASDPAKQVSIHNYTIWVSVLVAIWTVASLLPLPEWFFSTIRHHEGVE MCLVDVPAVAEEFMSMFGLYPLLAFFGLPLFFASFYFWRAYDQCK KRGTKTQNLNRNQIRSKQVTVMLLSIAHISALLWLPEWVAWLWVWH LKAAGPAPPQGFIALSQVLMFSISSANPLIFLVMSEEFREGLKGWVK WMITKKPPTVSESQETPAGNSEGLPDKVPSPEPASPEKEKPPSSPS GKGKTEKAEIPILPDVEQFWHERDTVPSVQDNDPIPWEHEDQETGE GVK
33	GPR161	MSLNSSLSCRKELSNLTEEKGEGGVITQFIAIIVITIFVCLGNLVIVV TLYKKSYYLLTSLNKFVFSLTLSNLLSVLVLPFVVTSSIRREWIFGVV WCNFSALLYLLISSASMLTLGVIAIDRYAVLYPMVYPMKITGNRA VMALVYIWLHSLIGCLPPLFGWSSVEFEFKWMCVAAWHREPGYT AFWQIWCALFPFLVMLVCYGFIFRVARVKARKVHCGTVVIVEEDA QRTGRKNSSTSTSSSGSRRNAFQGVVYSANQCKALITLVVLGAFM VTWGPYMVVIASEALWGKSSVSPSLETWATWLSFASAVCHPLIYGL WNKTVRKELLGMCFGDRYYREPFVQRQRTSRLFSISNRITDGLSPH LTALMAGGQPLGHSSSTGDTGFSCSQDSGTDMMLEDYTSDDNPPS HCTCPPKRRSSVTFEDEVQIKEAAKNSILHVKAHVHKSLSYAAASL AKAIEAEAKINLFGEEALPGVLVTARTVPGGGFGGRRGSRTLVSQR LQLQSIEEGDVLAEEQR
34	GPR17	MSKRWWAGSRKPPREMLKLSGSDSSQSMNGLEVAPPGLITNFSLA TAEQCGQETPLENMLFASFYLLDFILALVGNTLALWLFIRDHKSOTP ANVFLMHLAVADLSCVLVLPTRLVYHFSGNHWPFGIACRLTGFLF YLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLWV VVAVAMAPLLVSPQTVQTNHTVVCLQLYREKASHHALVSLAVAFT FPFITTVTCYLLIIRSLRQGLRVEKRLKTKAVRMIAIVLAIFLVCFVPY HVNRSVYVLHYRSHGASCATQRILALANRITSCLTSLNGALDPIMYF FVAEKFRHALCNLLCGKRLKGPPPSFEGKTNESSLSAKSEL
35	GPR183	MDIQMANNFTPPSATPQGNDCDLYAHHSTARIVMPLHYSLVFIIGLV GNLLALVVIVQNRKKINSTTLYSTNLVISDILFTALPTRIAYYAMGF DWRIGDALCRITALVFYINTYAGVNFMTCLSIDRFIAVVHPLRYNKI KRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITCMEYPNFEET KSLPWILLGACFIGYVLPILILICYSQICCKLFRITAKQNPLTEKSGVN KKALNTIILIVVFLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQIS LHFTVCLMNFNCCMDPFIYFFACKGYKRKVMRMLKRQVSVSISA VKSAPPEENSREMTETQMMIHSKSSNGK
36	CRTH2	MSANATLKPLCPILQMSRLQSHSNTSIRYIDHAAVLLHGLASLLGL VENGVILFVVGCRMRTQVTTWVHLALSDDLASASLPFFTYFLAV

		GHSWELGTTFCKLHSSIFFLNMFASGFLLSAISLDRCLQVVRPVWAQ NHRTVAAAHKVCLVLWALAVLNTVPYFVFRDTISRLDGRIMCYYN VLLLNPGPDRDATCNSRQVALAVSKFLLAFLVPLAIIASSHAAVSLR LQHRGRRRPGRFVRLVAAVVAAFALCWGPYHVFSLEARAHANPG LRPLVWRGLPFVTSLAFFNSVANPVLVLTCPDMLRKLRRSLRTVL ESVLVDDSELGGAGSSRRRTSSTARSASPLALCSRPEEPRGPARLL GWLLGSCAASPQTGPLNRALSSTSS
37	5-HT4	MDKLDANVSSEEGFGSVEKVLLTFLSTVILMAILGNLLVMVAVC WDRQLRKIKTNYFIVSLAFADLLVSVLVMPFGAIELVQDIWIYGEVF CLVRTSLDVLLTTASIFHLCCISLDRYYAICCCQPLVYRNKMTPLRIAL MLGGCWVPTFISFLPIMQGWNNIGIIDLIEKRKFNQNSNSTYCVFM VNKPYAITCSVVAFYIPFLLMVLAYYRIYVTAKEHAHQIQLQRAG ASSESRPQSADQHSTHRMRTETKAAKTLCIIMGCFCLCWAPFFVTNI VDPFIDYTVPGQVWTAFLWLGYINSGLNPFYAFLNKSFRRAFLIIL CCDDERYRRPSILGQTPVCSTTTINGSTHVLRLDAVECGGQWESQCH PPATSPLVAAQPSDT
38	5-HT6	MVPEPGPTANSTPAWGAGPPSAPGGSGWVAAALCVVIALTAAANS LLIALICTQPALRNTSNFFLVSLFTSDLMVGLVVMPPAMLNALYGR WVLARGLCLLWTAFDVMCCSASILNCLISLDRYLLILSPLRYKLR MTPLRALALVLGAWSLAALASFLPLLGWHELGHARPPVPGQCRL LASLPFVLVASGLTFFLPSGAICFTYCRILLAARKQAVQVASLTTGM ASQASETLQVPRTPRPGVESADSRRLATKHSRKALKASLTGLILLGM FFVTWLPFFVANIVQAVCDCISPGLFDVLTWLGYCNSTMNPIIYPLF MRDFKRALGRFLPCPRCPRERQASLASPSLRTSHSGPRPGLSLQQVL PLPLPPDSDSDSADAGSGGSSGLRLTAQLLLPGEATQDPPLPTRAAAA VNFFNIDPAEPELRPHPLGIPTN
39	CB2	MEECWVTEIANGSKDGLDSNPMKDYMILSGPQKTAVAVLCTLLGL LSALENVAVLYLILSSHQLRRKPSYLFIGSLAGADFLASVVFACSFV NFHVHGVDSKAVFLLKIGSVTMTFTASVGSLLLTAIDRYLCLRYPP SYKALLTRGRALVTLGIMWVLSALVSYLPLMGWTCPCRPCSELFPL IPNDYLLSWLLFIAFLFSGIITYTYGHVLWKAHQHVASLSGHQDRQVP GMARMRLDVRLAKTLGLVLAVLLICWFPVLALMAHSLATTLSDQV KKAFAFCSMLCLINSMVNPVIYALRSGEIRSSAHHCIAHWKCKVRG LGSEAKEEAPRSSVTETeadgKITPWPDSRDLDLSDC
40	Histamine-3	MERAPPDGPLNASGALAGEAAAAGGARGFSAAWTAVLAALMALL IVATVLGNALVMLAFVADSSLRTQNNFFLLNLAISDFLVGAFCIPLY VPYVLTGRWTFGRGLCKLWLVVDYLLCTSSAFNIVLISYDRFLSVT RAVSYRAQQGDRRAVRKMMLVWVLAFLLYGPAILSWEYLSGGSS IPEGHCYAEFFYNWYFLITASTLEFFTPLSVTFFNLSIYLNQRTRL RLDGAREAAGPEPPPEAQSPPPPPGCWGCWQKGHGEAMPLHRYG VGEAAVGAEAGEATLGGGGGGGSVASPTSSSGSSSRGTERPRSLKR GSKPSASSASLEKRMKMVSQSFTQRFRLSRDRKVAKSLAVIVSIFGL CWAPYTLLMIIRAACHGHCVPDYWYETSFWLLWANSVNPVLYPL CHHSFRRRAFTKLLCPQKLKIQPHSSLEHCWK
41	VPAC-1 or VIPR1	ARLQEECDYVQMIQVQHKQCLEEAQLENETIGCSKMWDNLTCPWA TPRGQVVVLACPLIFKLFSSIQGRNVSRCTDEGWTHLEPGPYPIAC GLDDKAASLDEQQTMFYGSVKTYTIGYGLSLATLLVATAILSLFR KLHCTRNYIHMHLFISFILRAAAVFIKDLALFDSGESDQCSEGSVGC KAAMVFFQYCVMANFFWLLVEGLYLYTLLAVSFFSERKYFWGYILI GWGVPSTFTMVWTIARIHFEDYGCWDTINSSLWWIKGPILTSILVN FILFICIIRILLQKLRRPDIRKSDSSPYSLARSTLLLIPLFGVHYIMFAF

		FPDNFKPEVKMVFELVVGSFQGFVVAILYCFLNGEVQAE LRKWR RWHLQGVLGWNPKYRHPSSGGSNGATCSTQVSMLTRVSPGARRSSS FQAEVSLV
42	GIPR	RAETGSKGQTAGELYQRWERYRRECQETLAAAEPSSGLACNGSFD MYVCWDYAAPNATARASCPWYLPWHHHVAAGFVLRQCGSDGQW GLWRDHTQCENPEKNEAFLDQRLILERLQVMYTVGYSLSLATLLA LLILSLFRRLHCTRNYIHINLFTS FMLRAAAILSRDRLLPRPGPYLGD QALALWNQALAAACRTAQIVTQYCVGANYTWLLVEGVYLHSLVL VGGSEEGHFRYYLLLGWGPALFVIPWVIVRYLYENTQCWERNEV KAIWWIIRTPILMTILINFLIFIRILGILLSKLRTRQMRCRDYRLRLARS TLTLVPLLGVHEVVFAVTEEQARGALRFAKLGFELSSSQGFLVS VLYCFINKEVQSEIRRGWHHCRLRRSLGEEQRQLPERAFRALPSGSG PGEVPTSRGLSSGTLPGPGNEASRELESYC
43	5-HT1B GPCR	MEEPGAQCAPPAPAGSETWVPQANLSSAPSQNC SAKDYIYQDSISLP WKVLLVMLLALITLATTLSNAFVIATVYRTRKLHTPANLYIASLAVT DLLVSILVMPISTMYTVTGRWTLGQVVCDFWLSSDITCCTASILHLC VIALDRYWAITDAVEYS AKRTPKRAAVMIALVWVFSISISLPPFFWR QAKAEVSECVVNTDHILYTVYSTVGAFYFPTLLLIALLYGRIYVEA RSRILKQTPNRTGKRLTRAQLITDSPGSTSSVTSINSRVPDVPSESGSP VYVNQVKVRVSDALLEKKKLMAARERKATKTLGILGAFIVCWLPF FIISLVPICKDACWFHLAIFDFFT WLGYLNSLINPIHYTMSNEDFKQ AFHKLIRFKCTS
44	CCR7	QDEVTDYIGDNTTVDYTLFESLCSKKDVRNFKAWFLPIMYSIICFV GLLGNGLVVLTYYIFKRLKTM TDYLLNLAVADILFLLTLPFWAYS AAKSWVFGVHFCKLIFAIYKMSFFSGM LLLLCISIDRYVAIVQAVSA HRHRARVLLISKLS CVGIWILATVLSIPELLYSDLQRSSEQAMRCSL ITEHVEAFITIQVAQMVIGFLVPLLAMSFCYLVIIRTLLQARNFERNK AIKVIIAVVVVFIVFQLPYNGV VLAQTVANFNITSSTCELSKQLNIA DVTYSLACVRCCVNPFLYAFIGVKFRNDL FKLFDLGLCSQEQLRQ WSSCRHIRRSSMSVEAETTTTFSP
45	CXCR5	MNYPLTLEMDLENLEDLFWELDRLDNYNDTSLVENHLC PATEGPL MASFKAVFVPVAYSLIFLLGVIGNVLVLVILERHRQTRSSTETFLFHL AVADLLL VFILPFAVAEGSVGWVLTGTF LCKTVIALHKVNFYCSSLL ACIAVDRYLAIVHAVHAYRHRRLSIHITCGTIWL VGFLALPEILFA KVSQGHNNSLPRCTFSQENQAETHAWFTSRFLYHVAGFLLPMLV MGWCYVG VVHRLRQAQRRPQRQKAVRVAILVTSIFFLCWSPYHIVI FLDTLARLKAVDNTCKLNGSLPVAITMCEFLGLAHCCLNPMLYTFA GVKFRSDLSRLLTKLGCTGPASLCQLFPSWRRSSLS ESENATSLTTF
46	GPR119	MESSFSFGVILAVLASLIATNTLVAVAVLLLIHKNDGVSLCFTLNL VADTLIGVAISGLLTDQLSSPSRPTQKTLCSLRMAFVTSSAAASVLT VMLITFDRYLAIKQPFYRLKIMSGFVAGACIAGLWLVSYLIGFLPLGI PMFQQTAYKGQCSFFAVFHPHFVLT LSCVGFFPAMLLFVFFYCDML KIASMHSQQIRKMEHAGAMAGGYRSPRTPSDFKALRTVSVLIGSFA LSWTPFLITGIVQVACQECHLYLV LERYLWLLGVGNSLLNPLIYAY WQKEVRLQLYHMALGVKKVLT SFLFLSARNCGPERPRESSCHIVT ISSSEFDG
47	GPR55	MSQQNTSGDCLFDGVNELMKT LQFAVHIPTFVLGLLNLLAIHGFS TFLKNRWPDYAATSIYMINLAVFDLLL VLSLPFKMVLVSQVQSPFPSL CTLVECLYFVSMYGSVFTICFISMDRFLAIRYPLL VSHLRSPRKIFGIC CTIWVLVWTGSIPIYSFHGKVEKYMCFHNMSDDTWSAKVFFPLEVF GFLPMGIMGFCCSRSIHILLGRRDHTQDWVQQKACIYSIAASLAVF

		VVSFLPVHLGFFLQFLVRNSFIVECRAKQSSISFFLQLSMCFNSVNCCL DVFCYYFVIKEFRMNIRAHPSRVQLVLQDTTISRG
--	--	--

[0072] Provided herein are scaffolds comprising GPCR binding domains, wherein the sequences of the GPCR binding domains support interaction with at least one GPCR. The sequence may be homologous or identical to a sequence of a GPCR ligand. In some instances, the GPCR binding domain sequence comprises at least or about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47. In some instances, the GPCR binding domain sequence comprises at least or about 95% homology to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47. In some instances, the GPCR binding domain sequence comprises at least or about 97% homology to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47. In some instances, the GPCR binding domain sequence comprises at least or about 99% homology to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47. In some instances, the GPCR binding domain sequence comprises at least or about 100% homology to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47. In some instances, the GPCR binding domain sequence comprises at least a portion having at least or about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, or more than 400 amino acids of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47.

[0073] Libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains may bind to one or more GPCRs. In some instances, the scaffolds comprising GPCR binding domains binds to a single GPCR. In some instances, the scaffolds comprising GPCR binding domains binds to GPCRs in a same family or class. In some instances, the scaffolds comprising GPCR binding domains bind to multiple GPCRs. For example, the scaffolds are multimeric and comprise at least 2 scaffolds. In some instances, the multimeric scaffolds comprise at least or about 3, 4, 5, 6, 7, 8, or more than 8 scaffolds. In some instances, the multimeric scaffolds comprise at least 2 scaffolds linked by, for example, a dimerization domain,

an amino acid linker, a disulfide bond, a chemical crosslink, or any other linker known in the art. In some instances, the multimeric scaffolds bind to the same GPCRs or different GPCRs.

[0074] Provided herein are GPCR binding libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains comprise variation in domain type, domain length, or residue variation. In some instances, the domain is a region in the scaffold comprising the GPCR binding domains. For example, the region is the VH, CDR-H3, or VL domain. In some instances, the domain is the GPCR binding domain.

[0075] Methods described herein provide for synthesis of a GPCR binding library of nucleic acids each encoding for a predetermined variant of at least one predetermined reference nucleic acid sequence. In some cases, the predetermined reference sequence is a nucleic acid sequence encoding for a protein, and the variant library comprises sequences encoding for variation of at least a single codon such that a plurality of different variants of a single residue in the subsequent protein encoded by the synthesized nucleic acid are generated by standard translation processes. In some instances, the GPCR binding library comprises varied nucleic acids collectively encoding variations at multiple positions. In some instances, the variant library comprises sequences encoding for variation of at least a single codon of a VH, CDR-H3, or VL domain. In some instances, the variant library comprises sequences encoding for variation of at least a single codon in a GPCR binding domain. For example, at least one single codon of a GPCR binding domain as listed in **Table 2** is varied. In some instances, the variant library comprises sequences encoding for variation of multiple codons of a VH, CDR-H3, or VL domain. In some instances, the variant library comprises sequences encoding for variation of multiple codons in a GPCR binding domain. An exemplary number of codons for variation include, but are not limited to, at least or about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 225, 250, 275, 300, or more than 300 codons.

[0076] Methods described herein provide for synthesis of a GPCR binding library of nucleic acids each encoding for a predetermined variant of at least one predetermined reference nucleic acid sequence, wherein the GPCR binding library comprises sequences encoding for variation of length of a domain. In some instances, the domain is VH, CDR-H3, or VL domain. In some instances, the domain is the GPCR binding domain. In some instances, the library comprises sequences encoding for variation of length of at least or about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 225, 250, 275, 300, or more than 300 codons less as compared to a predetermined reference sequence. In some instances, the library comprises sequences encoding for variation of length of at least or about 1, 5, 10, 15, 20, 25, 30,

35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, or more than 300 codons more as compared to a predetermined reference sequence.

[0077] Provided herein are GPCR binding libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains, wherein the GPCR binding libraries are synthesized with various numbers of fragments. In some instances, the fragments comprise the VH, CDR-H3, or VL domain. In some instances, the GPCR binding libraries are synthesized with at least or about 2 fragments, 3 fragments, 4 fragments, 5 fragments, or more than 5 fragments. The length of each of the nucleic acid fragments or average length of the nucleic acids synthesized may be at least or about 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, or more than 600 base pairs. In some instances, the length is about 50 to 600, 75 to 575, 100 to 550, 125 to 525, 150 to 500, 175 to 475, 200 to 450, 225 to 425, 250 to 400, 275 to 375, or 300 to 350 base pairs.

[0078] GPCR binding libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains as described herein comprise various lengths of amino acids when translated. In some instances, the length of each of the amino acid fragments or average length of the amino acid synthesized may be at least or about 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, or more than 150 amino acids. In some instances, the length of the amino acid is about 15 to 150, 20 to 145, 25 to 140, 30 to 135, 35 to 130, 40 to 125, 45 to 120, 50 to 115, 55 to 110, 60 to 110, 65 to 105, 70 to 100, or 75 to 95 amino acids. In some instances, the length of the amino acid is about 22 to about 75 amino acids.

[0079] GPCR binding libraries comprising de novo synthesized variant sequences encoding for scaffolds comprising GPCR binding domains comprise a number of variant sequences. In some instances, a number of variant sequences is de novo synthesized for a CDR-H1, CDR-H2, CDR-H3, CDR-L1, CDR-L2, CDR-L3, VL, VH, or a combination thereof. In some instances, a number of variant sequences is de novo synthesized for framework element 1 (FW1), framework element 2 (FW2), framework element 3 (FW3), or framework element 4 (FW4). In some instances, a number of variant sequences is de novo synthesized for a GPCR binding domain. For example, the number of variant sequences is about 1 to about 10 sequences for the VH domain, about 10^8 sequences for the GPCR binding domain, and about 1 to about 44 sequences for the VK domain. See FIG. 2. The number of variant sequences may be at least or about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, or more than 500 sequences. In some

instances, the number of variant sequences is about 10 to 300, 25 to 275, 50 to 250, 75 to 225, 100 to 200, or 125 to 150 sequences.

[0080] GPCR binding libraries comprising de novo synthesized variant sequences encoding for scaffolds comprising GPCR binding domains comprise improved diversity. For example, variants are generated by placing GPCR binding domain variants in immunoglobulin scaffold variants comprising N-terminal CDR-H3 variations and C-terminal CDR-H3 variations. In some instances, variants include affinity maturation variants. Alternatively or in combination, variants include variants in other regions of the immunoglobulin including, but not limited to, CDR-H1, CDR-H2, CDR-L1, CDR-L2, and CDR-L3. In some instances, the number of variants of the GPCR binding libraries is least or about 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , or more than 10^{10} non-identical sequences. For example, a library comprising about 10 variant sequences for a VH region, about 237 variant sequences for a CDR-H3 region, and about 43 variant sequences for a VL and CDR-L3 region comprises 10^5 non-identical sequences ($10 \times 237 \times 43$). See FIGS. 4A-4B.

[0081] Following synthesis of GPCR binding libraries comprising nucleic acids encoding scaffolds comprising GPCR binding domains, libraries may be used for screening and analysis. For example, libraries are assayed for library displayability and panning. In some instances, displayability is assayed using a selectable tag. Exemplary tags include, but are not limited to, a radioactive label, a fluorescent label, an enzyme, a chemiluminescent tag, a colorimetric tag, an affinity tag or other labels or tags that are known in the art. In some instances, the tag is histidine, polyhistidine, myc, hemagglutinin (HA), or FLAG. For example as seen in FIG. 3, the GPCR binding libraries comprises nucleic acids encoding scaffolds comprising GPCR binding domains with multiple tags such as GFP, FLAG, and Lucy as well as a DNA barcode. In some instances, libraries are assayed by sequencing using various methods including, but not limited to, single-molecule real-time (SMRT) sequencing, Polony sequencing, sequencing by ligation, reversible terminator sequencing, proton detection sequencing, ion semiconductor sequencing, nanopore sequencing, electronic sequencing, pyrosequencing, Maxam-Gilbert sequencing, chain termination (e.g., Sanger) sequencing, +S sequencing, or sequencing by synthesis.

[0082] *Expression Systems*

[0083] Provided herein are libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains, wherein the libraries have improved specificity, stability, expression, folding, or downstream activity. In some instances, libraries described herein are used for screening and analysis.

[0084] Provided herein are libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains, wherein the nucleic acid libraries are used for screening and analysis. In some instances, screening and analysis comprises *in vitro*, *in vivo*, or *ex vivo* assays. Cells for screening include primary cells taken from living subjects or cell lines. Cells may be from prokaryotes (e.g., bacteria and fungi) or eukaryotes (e.g., animals and plants). Exemplary animal cells include, without limitation, those from a mouse, rabbit, primate, and insect. In some instances, cells for screening include a cell line including, but not limited to, Chinese Hamster Ovary (CHO) cell line, human embryonic kidney (HEK) cell line, or baby hamster kidney (BHK) cell line. In some instances, nucleic acid libraries described herein may also be delivered to a multicellular organism. Exemplary multicellular organisms include, without limitation, a plant, a mouse, rabbit, primate, and insect.

[0085] Nucleic acid libraries described herein may be screened for various pharmacological or pharmacokinetic properties. In some instances, the libraries are screened using *in vitro* assays, *in vivo* assays, or *ex vivo* assays. For example, *in vitro* pharmacological or pharmacokinetic properties that are screened include, but are not limited to, binding affinity, binding specificity, and binding avidity. Exemplary *in vivo* pharmacological or pharmacokinetic properties of libraries described herein that are screened include, but are not limited to, therapeutic efficacy, activity, preclinical toxicity properties, clinical efficacy properties, clinical toxicity properties, immunogenicity, potency, and clinical safety properties.

[0086] Provided herein are nucleic acid libraries, wherein the nucleic acid libraries may be expressed in a vector. Expression vectors for inserting nucleic acid libraries disclosed herein may comprise eukaryotic or prokaryotic expression vectors. Exemplary expression vectors include, without limitation, mammalian expression vectors: pSF-CMV-NEO-NH2-PPT-3XFLAG, pSF-CMV-NEO-COOH-3XFLAG, pSF-CMV-PURO-NH2-GST-TEV, pSF-OXB20-COOH-TEV-FLAG(R)-6His, pCEP4 pDEST27, pSF-CMV-Ub-KrYFP, pSF-CMV-FMDV-daGFP, pEF1a-mCherry-N1 Vector, pEF1a-tdTomato Vector, pSF-CMV-FMDV-Hygro, pSF-CMV-PGK-Puro, pMCP-tag(m), and pSF-CMV-PURO-NH2-CMYC; bacterial expression vectors: pSF-OXB20-BetaGal, pSF-OXB20-Fluc, pSF-OXB20, and pSF-Tac; plant expression vectors: pRI 101-AN DNA and pCambia2301; and yeast expression vectors: pTYB21 and pKLAC2, and insect vectors: pAc5.1/V5-His A and pDEST8. In some instances, the vector is pcDNA3 or pcDNA3.1.

[0087] Described herein are nucleic acid libraries that are expressed in a vector to generate a construct comprising a scaffold comprising sequences of GPCR binding domains. In some instances, a size of the construct varies. In some instances, the construct comprises at least or

about 500, 600, 700, 800, 900, 1000, 1100, 1300, 1400, 1500, 1600, 1700, 1800, 2000, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, 5000, 6000, 7000, 8000, 9000, 10000, or more than 10000 bases. In some instances, a the construct comprises a range of about 300 to 1,000, 300 to 2,000, 300 to 3,000, 300 to 4,000, 300 to 5,000, 300 to 6,000, 300 to 7,000, 300 to 8,000, 300 to 9,000, 300 to 10,000, 1,000 to 2,000, 1,000 to 3,000, 1,000 to 4,000, 1,000 to 5,000, 1,000 to 6,000, 1,000 to 7,000, 1,000 to 8,000, 1,000 to 9,000, 1,000 to 10,000, 2,000 to 3,000, 2,000 to 4,000, 2,000 to 5,000, 2,000 to 6,000, 2,000 to 7,000, 2,000 to 8,000, 2,000 to 9,000, 2,000 to 10,000, 3,000 to 4,000, 3,000 to 5,000, 3,000 to 6,000, 3,000 to 7,000, 3,000 to 8,000, 3,000 to 9,000, 3,000 to 10,000, 4,000 to 5,000, 4,000 to 6,000, 4,000 to 7,000, 4,000 to 8,000, 4,000 to 9,000, 4,000 to 10,000, 5,000 to 6,000, 5,000 to 7,000, 5,000 to 8,000, 5,000 to 9,000, 5,000 to 10,000, 6,000 to 7,000, 6,000 to 8,000, 6,000 to 9,000, 6,000 to 10,000, 7,000 to 8,000, 7,000 to 9,000, 7,000 to 10,000, 8,000 to 9,000, 8,000 to 10,000, or 9,000 to 10,000 bases.

[0088] Provided herein are libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains, wherein the nucleic acid libraries are expressed in a cell. In some instances, the libraries are synthesized to express a reporter gene. Exemplary reporter genes include, but are not limited to, acetohydroxyacid synthase (AHAS), alkaline phosphatase (AP), beta galactosidase (LacZ), beta glucuronidase (GUS), chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), red fluorescent protein (RFP), yellow fluorescent protein (YFP), cyan fluorescent protein (CFP), cerulean fluorescent protein, citrine fluorescent protein, orange fluorescent protein, cherry fluorescent protein, turquoise fluorescent protein, blue fluorescent protein, horseradish peroxidase (HRP), luciferase (Luc), nopaline synthase (NOS), octopine synthase (OCS), luciferase, and derivatives thereof. Methods to determine modulation of a reporter gene are well known in the art, and include, but are not limited to, fluorometric methods (*e.g.* fluorescence spectroscopy, Fluorescence Activated Cell Sorting (FACS), fluorescence microscopy), and antibiotic resistance determination.

[0089] Diseases and Disorders

[0090] Provided herein are GPCR binding libraries comprising nucleic acids encoding for scaffolds comprising GPCR binding domains may have therapeutic effects. In some instances, the GPCR binding libraries result in protein when translated that is used to treat a disease or disorder. In some instances, the protein is an immunoglobulin. In some instances, the protein is a peptidomimetic. Exemplary diseases include, but are not limited to, cancer, inflammatory diseases or disorders, a metabolic disease or disorder, a cardiovascular disease or disorder, a respiratory disease or disorder, pain, a digestive disease or disorder, a reproductive disease or

disorder, an endocrine disease or disorder, or a neurological disease or disorder. In some instances, the cancer is a solid cancer or a hematologic cancer. In some instances, an inhibitor of GPCR glucagon like peptide 1 receptor (GLP1R) as described herein is used for treatment of a metabolic disorder. In some instances, an inhibitor of GPCR GLP1R as described herein is used for treatment of weight gain (or for inducing weight loss), treatment of obesity, or treatment of Type II diabetes. In some instances, the subject is a mammal. In some instances, the subject is a mouse, rabbit, dog, or human. Subjects treated by methods described herein may be infants, adults, or children. Pharmaceutical compositions comprising antibodies or antibody fragments as described herein may be administered intravenously or subcutaneously. In some instances, a pharmaceutical composition comprises an antibody or antibody fragment described herein comprising a CDR-H3 comprising a sequence of any one of SEQ ID NOS: 2420 to 2436. In further instances, the pharmaceutical composition is used for treatment of a metabolic disorder.

[0091] Variant Libraries

[0092] *Codon variation*

[0093] Variant nucleic acid libraries described herein may comprise a plurality of nucleic acids, wherein each nucleic acid encodes for a variant codon sequence compared to a reference nucleic acid sequence. In some instances, each nucleic acid of a first nucleic acid population contains a variant at a single variant site. In some instances, the first nucleic acid population contains a plurality of variants at a single variant site such that the first nucleic acid population contains more than one variant at the same variant site. The first nucleic acid population may comprise nucleic acids collectively encoding multiple codon variants at the same variant site. The first nucleic acid population may comprise nucleic acids collectively encoding up to 19 or more codons at the same position. The first nucleic acid population may comprise nucleic acids collectively encoding up to 60 variant triplets at the same position, or the first nucleic acid population may comprise nucleic acids collectively encoding up to 61 different triplets of codons at the same position. Each variant may encode for a codon that results in a different amino acid during translation. **Table 3** provides a listing of each codon possible (and the representative amino acid) for a variant site.

Table 3. List of codons and amino acids

Amino Acids	One letter code	Three letter code	Codons

Alanine	A	Ala	GCA	GCC	GCG	GCT		
Cysteine	C	Cys	TGC	TGT				
Aspartic acid	D	Asp	GAC	GAT				
Glutamic acid	E	Glu	GAA	GAG				
Phenylalanine	F	Phe	TTC	TTT				
Glycine	G	Gly	GGA	GGC	GGG	GGT		
Histidine	H	His	CAC	CAT				
Isoleucine	I	Iso	ATA	ATC	ATT			
Lysine	K	Lys	AAA	AAG				
Leucine	L	Leu	TTA	TTG	CTA	CTC	CTG	CTT
Methionine	M	Met	ATG					
Asparagine	N	Asn	AAC	AAT				
Proline	P	Pro	CCA	CCC	CCG	CCT		
Glutamine	Q	Gln	CAA	CAG				
Arginine	R	Arg	AGA	AGG	CGA	CGC	CGG	CGT
Serine	S	Ser	AGC	AGT	TCA	TCC	TCG	TCT
Threonine	T	Thr	ACA	ACC	ACG	ACT		
Valine	V	Val	GTA	GTC	GTG	GTT		
Tryptophan	W	Trp	TGG					
Tyrosine	Y	Tyr	TAC	TAT				

[0094] A nucleic acid population may comprise varied nucleic acids collectively encoding up to 20 codon variations at multiple positions. In such cases, each nucleic acid in the population comprises variation for codons at more than one position in the same nucleic acid. In some

instances, each nucleic acid in the population comprises variation for codons at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more codons in a single nucleic acid. In some instances, each variant long nucleic acid comprises variation for codons at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more codons in a single long nucleic acid. In some instances, the variant nucleic acid population comprises variation for codons at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more codons in a single nucleic acid. In some instances, the variant nucleic acid population comprises variation for codons in at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more codons in a single long nucleic acid.

[0095] Highly Parallel Nucleic Acid Synthesis

[0096] Provided herein is a platform approach utilizing miniaturization, parallelization, and vertical integration of the end-to-end process from polynucleotide synthesis to gene assembly within nanowells on silicon to create a revolutionary synthesis platform. Devices described herein provide, with the same footprint as a 96-well plate, a silicon synthesis platform is capable of increasing throughput by a factor of up to 1,000 or more compared to traditional synthesis methods, with production of up to approximately 1,000,000 or more polynucleotides, or 10,000 or more genes in a single highly-parallelized run.

[0097] With the advent of next-generation sequencing, high resolution genomic data has become an important factor for studies that delve into the biological roles of various genes in both normal biology and disease pathogenesis. At the core of this research is the central dogma of molecular biology and the concept of “residue-by-residue transfer of sequential information.” Genomic information encoded in the DNA is transcribed into a message that is then translated into the protein that is the active product within a given biological pathway.

[0098] Another exciting area of study is on the discovery, development and manufacturing of therapeutic molecules focused on a highly-specific cellular target. High diversity DNA sequence libraries are at the core of development pipelines for targeted therapeutics. Gene mutants are used to express proteins in a design, build, and test protein engineering cycle that ideally culminates in an optimized gene for high expression of a protein with high affinity for its therapeutic target. As an example, consider the binding pocket of a receptor. The ability to test all sequence permutations of all residues within the binding pocket simultaneously will allow for a thorough exploration, increasing chances of success. Saturation mutagenesis, in which a researcher attempts to generate all possible mutations at a specific site within the receptor, represents one approach to this development challenge. Though costly and time and labor-intensive, it enables each variant to be introduced into each position. In contrast, combinatorial

mutagenesis, where a few selected positions or short stretch of DNA may be modified extensively, generates an incomplete repertoire of variants with biased representation.

[0099] To accelerate the drug development pipeline, a library with the desired variants available at the intended frequency in the right position available for testing—in other words, a precision library, enables reduced costs as well as turnaround time for screening. Provided herein are methods for synthesizing nucleic acid synthetic variant libraries which provide for precise introduction of each intended variant at the desired frequency. To the end user, this translates to the ability to not only thoroughly sample sequence space but also be able to query these hypotheses in an efficient manner, reducing cost and screening time. Genome-wide editing can elucidate important pathways, libraries where each variant and sequence permutation can be tested for optimal functionality, and thousands of genes can be used to reconstruct entire pathways and genomes to re-engineer biological systems for drug discovery.

[00100] In a first example, a drug itself can be optimized using methods described herein. For example, to improve a specified function of an antibody, a variant polynucleotide library encoding for a portion of the antibody is designed and synthesized. A variant nucleic acid library for the antibody can then be generated by processes described herein (e.g., PCR mutagenesis followed by insertion into a vector). The antibody is then expressed in a production cell line and screened for enhanced activity. Example screens include examining modulation in binding affinity to an antigen, stability, or effector function (e.g., ADCC, complement, or apoptosis). Exemplary regions to optimize the antibody include, without limitation, the Fc region, Fab region, variable region of the Fab region, constant region of the Fab region, variable domain of the heavy chain or light chain (V_H or V_L), and specific complementarity-determining regions (CDRs) of V_H or V_L .

[00101] Nucleic acid libraries synthesized by methods described herein may be expressed in various cells associated with a disease state. Cells associated with a disease state include cell lines, tissue samples, primary cells from a subject, cultured cells expanded from a subject, or cells in a model system. Exemplary model systems include, without limitation, plant and animal models of a disease state.

[00102] To identify a variant molecule associated with prevention, reduction or treatment of a disease state, a variant nucleic acid library described herein is expressed in a cell associated with a disease state, or one in which a cell a disease state can be induced. In some instances, an agent is used to induce a disease state in cells. Exemplary tools for disease state induction include, without limitation, a Cre/Lox recombination system, LPS inflammation induction, and streptozotocin to induce hypoglycemia. The cells associated with a disease state may be cells

from a model system or cultured cells, as well as cells from a subject having a particular disease condition. Exemplary disease conditions include a bacterial, fungal, viral, autoimmune, or proliferative disorder (e.g., cancer). In some instances, the variant nucleic acid library is expressed in the model system, cell line, or primary cells derived from a subject, and screened for changes in at least one cellular activity. Exemplary cellular activities include, without limitation, proliferation, cycle progression, cell death, adhesion, migration, reproduction, cell signaling, energy production, oxygen utilization, metabolic activity, and aging, response to free radical damage, or any combination thereof.

[00103] Substrates

[00104] Devices used as a surface for polynucleotide synthesis may be in the form of substrates which include, without limitation, homogenous array surfaces, patterned array surfaces, channels, beads, gels, and the like. Provided herein are substrates comprising a plurality of clusters, wherein each cluster comprises a plurality of loci that support the attachment and synthesis of polynucleotides. In some instances, substrates comprise a homogenous array surface. For example, the homogenous array surface is a homogenous plate. The term “locus” as used herein refers to a discrete region on a structure which provides support for polynucleotides encoding for a single predetermined sequence to extend from the surface. In some instances, a locus is on a two dimensional surface, *e.g.*, a substantially planar surface. In some instances, a locus is on a three-dimensional surface, *e.g.*, a well, microwell, channel, or post. In some instances, a surface of a locus comprises a material that is actively functionalized to attach to at least one nucleotide for polynucleotide synthesis, or preferably, a population of identical nucleotides for synthesis of a population of polynucleotides. In some instances, polynucleotide refers to a population of polynucleotides encoding for the same nucleic acid sequence. In some cases, a surface of a substrate is inclusive of one or a plurality of surfaces of a substrate. The average error rates for polynucleotides synthesized within a library described here using the systems and methods provided are often less than 1 in 1000, less than about 1 in 2000, less than about 1 in 3000 or less often without error correction.

[00105] Provided herein are surfaces that support the parallel synthesis of a plurality of polynucleotides having different predetermined sequences at addressable locations on a common support. In some instances, a substrate provides support for the synthesis of more than 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2,000; 5,000; 10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more non-identical polynucleotides. In some

cases, the surfaces provide support for the synthesis of more than 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2,000; 5,000; 10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more polynucleotides encoding for distinct sequences. In some instances, at least a portion of the polynucleotides have an identical sequence or are configured to be synthesized with an identical sequence. In some instances, the substrate provides a surface environment for the growth of polynucleotides having at least 80, 90, 100, 120, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 bases or more.

[00106] Provided herein are methods for polynucleotide synthesis on distinct loci of a substrate, wherein each locus supports the synthesis of a population of polynucleotides. In some cases, each locus supports the synthesis of a population of polynucleotides having a different sequence than a population of polynucleotides grown on another locus. In some instances, each polynucleotide sequence is synthesized with 1, 2, 3, 4, 5, 6, 7, 8, 9 or more redundancy across different loci within the same cluster of loci on a surface for polynucleotide synthesis. In some instances, the loci of a substrate are located within a plurality of clusters. In some instances, a substrate comprises at least 10, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000 or more clusters. In some instances, a substrate comprises more than 2,000; 5,000; 10,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,100,000; 1,200,000; 1,300,000; 1,400,000; 1,500,000; 1,600,000; 1,700,000; 1,800,000; 1,900,000; 2,000,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; or 10,000,000 or more distinct loci. In some instances, a substrate comprises about 10,000 distinct loci. The amount of loci within a single cluster is varied in different instances. In some cases, each cluster includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 300, 400, 500 or more loci. In some instances, each cluster includes about 50-500 loci. In some instances, each cluster includes about 100-200 loci. In some instances, each cluster includes about 100-150 loci. In some instances, each cluster includes about 109, 121, 130 or 137 loci. In some instances, each cluster includes about 19, 20, 61, 64 or more loci. Alternatively or in combination, polynucleotide synthesis occurs on a homogenous array surface.

[00107] In some instances, the number of distinct polynucleotides synthesized on a substrate is dependent on the number of distinct loci available in the substrate. In some instances, the

density of loci within a cluster or surface of a substrate is at least or about 1, 10, 25, 50, 65, 75, 100, 130, 150, 175, 200, 300, 400, 500, 1,000 or more loci per mm^2 . In some cases, a substrate comprises 10-500, 25-400, 50-500, 100-500, 150-500, 10-250, 50-250, 10-200, or 50-200 mm^2 . In some instances, the distance between the centers of two adjacent loci within a cluster or surface is from about 10-500, from about 10-200, or from about 10-100 μm . In some instances, the distance between two centers of adjacent loci is greater than about 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 μm . In some instances, the distance between the centers of two adjacent loci is less than about 200, 150, 100, 80, 70, 60, 50, 40, 30, 20 or 10 μm . In some instances, each locus has a width of about 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 μm . In some cases, each locus is has a width of about 0.5-100, 0.5-50, 10-75, or 0.5-50 μm .

[00108] In some instances, the density of clusters within a substrate is at least or about 1 cluster per 100 mm^2 , 1 cluster per 10 mm^2 , 1 cluster per 5 mm^2 , 1 cluster per 4 mm^2 , 1 cluster per 3 mm^2 , 1 cluster per 2 mm^2 , 1 cluster per 1 mm^2 , 2 clusters per 1 mm^2 , 3 clusters per 1 mm^2 , 4 clusters per 1 mm^2 , 5 clusters per 1 mm^2 , 10 clusters per 1 mm^2 , 50 clusters per 1 mm^2 or more. In some instances, a substrate comprises from about 1 cluster per 10 mm^2 to about 10 clusters per 1 mm^2 . In some instances, the distance between the centers of two adjacent clusters is at least or about 50, 100, 200, 500, 1000, 2000, or 5000 μm . In some cases, the distance between the centers of two adjacent clusters is between about 50-100, 50-200, 50-300, 50-500, and 100-2000 μm . In some cases, the distance between the centers of two adjacent clusters is between about 0.05-50, 0.05-10, 0.05-5, 0.05-4, 0.05-3, 0.05-2, 0.1-10, 0.2-10, 0.3-10, 0.4-10, 0.5-10, 0.5-5, or 0.5-2 mm. In some cases, each cluster has a cross section of about 0.5 to about 2, about 0.5 to about 1, or about 1 to about 2 mm. In some cases, each cluster has a cross section of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm. In some cases, each cluster has an interior cross section of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm.

[00109] In some instances, a substrate is about the size of a standard 96 well plate, for example between about 100 and about 200 mm by between about 50 and about 150 mm. In some instances, a substrate has a diameter less than or equal to about 1000, 500, 450, 400, 300, 250, 200, 150, 100 or 50 mm. In some instances, the diameter of a substrate is between about 25-1000, 25-800, 25-600, 25-500, 25-400, 25-300, or 25-200 mm. In some instances, a substrate has a planar surface area of at least about 100; 200; 500; 1,000; 2,000; 5,000; 10,000; 12,000; 15,000; 20,000; 30,000; 40,000; 50,000 mm^2 or more. In some instances, the thickness of a substrate is between about 50- 2000, 50- 1000, 100-1000, 200-1000, or 250-1000 mm.

[00110] *Surface materials*

[00111] Substrates, devices, and reactors provided herein are fabricated from any variety of materials suitable for the methods, compositions, and systems described herein. In certain instances, substrate materials are fabricated to exhibit a low level of nucleotide binding. In some instances, substrate materials are modified to generate distinct surfaces that exhibit a high level of nucleotide binding. In some instances, substrate materials are transparent to visible and/or UV light. In some instances, substrate materials are sufficiently conductive, *e.g.*, are able to form uniform electric fields across all or a portion of a substrate. In some instances, conductive materials are connected to an electric ground. In some instances, the substrate is heat conductive or insulated. In some instances, the materials are chemical resistant and heat resistant to support chemical or biochemical reactions, for example polynucleotide synthesis reaction processes. In some instances, a substrate comprises flexible materials. For flexible materials, materials can include, without limitation: nylon, both modified and unmodified, nitrocellulose, polypropylene, and the like. In some instances, a substrate comprises rigid materials. For rigid materials, materials can include, without limitation: glass; fused silica; silicon, plastics (for example polytetrafluoroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and the like); metals (for example, gold, platinum, and the like). The substrate, solid support or reactors can be fabricated from a material selected from the group consisting of silicon, polystyrene, agarose, dextran, cellulosic polymers, polyacrylamides, polydimethylsiloxane (PDMS), and glass. The substrates/solid supports or the microstructures, reactors therein may be manufactured with a combination of materials listed herein or any other suitable material known in the art.

[00112] Surface Architecture

[00113] Provided herein are substrates for the methods, compositions, and systems described herein, wherein the substrates have a surface architecture suitable for the methods, compositions, and systems described herein. In some instances, a substrate comprises raised and/or lowered features. One benefit of having such features is an increase in surface area to support polynucleotide synthesis. In some instances, a substrate having raised and/or lowered features is referred to as a three-dimensional substrate. In some cases, a three-dimensional substrate comprises one or more channels. In some cases, one or more loci comprise a channel. In some cases, the channels are accessible to reagent deposition via a deposition device such as a material deposition device. In some cases, reagents and/or fluids collect in a larger well in fluid communication one or more channels. For example, a substrate comprises a plurality of channels corresponding to a plurality of loci with a cluster, and the plurality of channels are in fluid communication with one well of the cluster. In some methods, a library of polynucleotides is synthesized in a plurality of loci of a cluster.

[00114] Provided herein are substrates for the methods, compositions, and systems described herein, wherein the substrates are configured for polynucleotide synthesis. In some instances, the structure is configured to allow for controlled flow and mass transfer paths for polynucleotide synthesis on a surface. In some instances, the configuration of a substrate allows for the controlled and even distribution of mass transfer paths, chemical exposure times, and/or wash efficacy during polynucleotide synthesis. In some instances, the configuration of a substrate allows for increased sweep efficiency, for example by providing sufficient volume for a growing polynucleotide such that the excluded volume by the growing polynucleotide does not take up more than 50, 45, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1%, or less of the initially available volume that is available or suitable for growing the polynucleotide. In some instances, a three-dimensional structure allows for managed flow of fluid to allow for the rapid exchange of chemical exposure.

[00115] Provided herein are substrates for the methods, compositions, and systems described herein, wherein the substrates comprise structures suitable for the methods, compositions, and systems described herein. In some instances, segregation is achieved by physical structure. In some instances, segregation is achieved by differential functionalization of the surface generating active and passive regions for polynucleotide synthesis. In some instances, differential functionalization is achieved by alternating the hydrophobicity across the substrate surface, thereby creating water contact angle effects that cause beading or wetting of the deposited reagents. Employing larger structures can decrease splashing and cross-contamination of distinct polynucleotide synthesis locations with reagents of the neighboring spots. In some cases, a device, such as a material deposition device, is used to deposit reagents to distinct polynucleotide synthesis locations. Substrates having three-dimensional features are configured in a manner that allows for the synthesis of a large number of polynucleotides (*e.g.*, more than about 10,000) with a low error rate (*e.g.*, less than about 1:500, 1:1000, 1:1500, 1:2,000, 1:3,000, 1:5,000, or 1:10,000). In some cases, a substrate comprises features with a density of about or greater than about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400 or 500 features per mm².

[00116] A well of a substrate may have the same or different width, height, and/or volume as another well of the substrate. A channel of a substrate may have the same or different width, height, and/or volume as another channel of the substrate. In some instances, the diameter of a cluster or the diameter of a well comprising a cluster, or both, is between about 0.05-50, 0.05-10, 0.05-5, 0.05-4, 0.05-3, 0.05-2, 0.05-1, 0.05-0.5, 0.05-0.1, 0.1-10, 0.2-10, 0.3-10, 0.4-10, 0.5-10, 0.5-5, or 0.5-2 mm. In some instances, the diameter of a cluster or well or both is less than or

about 5, 4, 3, 2, 1, 0.5, 0.1, 0.09, 0.08, 0.07, 0.06, or 0.05 mm. In some instances, the diameter of a cluster or well or both is between about 1.0 and 1.3 mm. In some instances, the diameter of a cluster or well, or both is about 1.150 mm. In some instances, the diameter of a cluster or well, or both is about 0.08 mm. The diameter of a cluster refers to clusters within a two-dimensional or three-dimensional substrate.

[00117] In some instances, the height of a well is from about 20-1000, 50-1000, 100- 1000, 200-1000, 300-1000, 400-1000, or 500-1000 μm . In some cases, the height of a well is less than about 1000, 900, 800, 700, or 600 μm .

[00118] In some instances, a substrate comprises a plurality of channels corresponding to a plurality of loci within a cluster, wherein the height or depth of a channel is 5-500, 5-400, 5-300, 5-200, 5-100, 5-50, or 10-50 μm . In some cases, the height of a channel is less than 100, 80, 60, 40, or 20 μm .

[00119] In some instances, the diameter of a channel, locus (*e.g.*, in a substantially planar substrate) or both channel and locus (*e.g.*, in a three-dimensional substrate wherein a locus corresponds to a channel) is from about 1-1000, 1-500, 1-200, 1-100, 5-100, or 10-100 μm , for example, about 90, 80, 70, 60, 50, 40, 30, 20 or 10 μm . In some instances, the diameter of a channel, locus, or both channel and locus is less than about 100, 90, 80, 70, 60, 50, 40, 30, 20 or 10 μm . In some instances, the distance between the center of two adjacent channels, loci, or channels and loci is from about 1-500, 1-200, 1-100, 5-200, 5-100, 5-50, or 5-30, for example, about 20 μm .

[00120] Surface Modifications

[00121] Provided herein are methods for polynucleotide synthesis on a surface, wherein the surface comprises various surface modifications. In some instances, the surface modifications are employed for the chemical and/or physical alteration of a surface by an additive or subtractive process to change one or more chemical and/or physical properties of a substrate surface or a selected site or region of a substrate surface. For example, surface modifications include, without limitation, (1) changing the wetting properties of a surface, (2) functionalizing a surface, *i.e.*, providing, modifying or substituting surface functional groups, (3) defunctionalizing a surface, *i.e.*, removing surface functional groups, (4) otherwise altering the chemical composition of a surface, *e.g.*, through etching, (5) increasing or decreasing surface roughness, (6) providing a coating on a surface, *e.g.*, a coating that exhibits wetting properties that are different from the wetting properties of the surface, and/or (7) depositing particulates on a surface.

[00122] In some cases, the addition of a chemical layer on top of a surface (referred to as adhesion promoter) facilitates structured patterning of loci on a surface of a substrate. Exemplary surfaces for application of adhesion promotion include, without limitation, glass, silicon, silicon dioxide and silicon nitride. In some cases, the adhesion promoter is a chemical with a high surface energy. In some instances, a second chemical layer is deposited on a surface of a substrate. In some cases, the second chemical layer has a low surface energy. In some cases, surface energy of a chemical layer coated on a surface supports localization of droplets on the surface. Depending on the patterning arrangement selected, the proximity of loci and/or area of fluid contact at the loci are alterable.

[00123] In some instances, a substrate surface, or resolved loci, onto which nucleic acids or other moieties are deposited, *e.g.*, for polynucleotide synthesis, are smooth or substantially planar (*e.g.*, two-dimensional) or have irregularities, such as raised or lowered features (*e.g.*, three-dimensional features). In some instances, a substrate surface is modified with one or more different layers of compounds. Such modification layers of interest include, without limitation, inorganic and organic layers such as metals, metal oxides, polymers, small organic molecules and the like.

[00124] In some instances, resolved loci of a substrate are functionalized with one or more moieties that increase and/or decrease surface energy. In some cases, a moiety is chemically inert. In some cases, a moiety is configured to support a desired chemical reaction, for example, one or more processes in a polynucleotide synthesis reaction. The surface energy, or hydrophobicity, of a surface is a factor for determining the affinity of a nucleotide to attach onto the surface. In some instances, a method for substrate functionalization comprises: (a) providing a substrate having a surface that comprises silicon dioxide; and (b) silanizing the surface using, a suitable silanizing agent described herein or otherwise known in the art, for example, an organofunctional alkoxy silane molecule. Methods and functionalizing agents are described in U.S. Patent No. 5474796, which is herein incorporated by reference in its entirety.

[00125] In some instances, a substrate surface is functionalized by contact with a derivatizing composition that contains a mixture of silanes, under reaction conditions effective to couple the silanes to the substrate surface, typically via reactive hydrophilic moieties present on the substrate surface. Silanization generally covers a surface through self-assembly with organofunctional alkoxy silane molecules. A variety of siloxane functionalizing reagents can further be used as currently known in the art, *e.g.*, for lowering or increasing surface energy. The organofunctional alkoxy silanes are classified according to their organic functions.

[00126] Polynucleotide Synthesis

[00127] Methods of the current disclosure for polynucleotide synthesis may include processes involving phosphoramidite chemistry. In some instances, polynucleotide synthesis comprises coupling a base with phosphoramidite. Polynucleotide synthesis may comprise coupling a base by deposition of phosphoramidite under coupling conditions, wherein the same base is optionally deposited with phosphoramidite more than once, i.e., double coupling. Polynucleotide synthesis may comprise capping of unreacted sites. In some instances, capping is optional.

Polynucleotide synthesis may also comprise oxidation or an oxidation step or oxidation steps.

Polynucleotide synthesis may comprise deblocking, detritylation, and sulfurization. In some instances, polynucleotide synthesis comprises either oxidation or sulfurization. In some instances, between one or each step during a polynucleotide synthesis reaction, the device is washed, for example, using tetrazole or acetonitrile. Time frames for any one step in a phosphoramidite synthesis method may be less than about 2 min, 1 min, 50 sec, 40 sec, 30 sec, 20 sec and 10 sec.

[00128] Polynucleotide synthesis using a phosphoramidite method may comprise a subsequent addition of a phosphoramidite building block (*e.g.*, nucleoside phosphoramidite) to a growing polynucleotide chain for the formation of a phosphite triester linkage. Phosphoramidite polynucleotide synthesis proceeds in the 3' to 5' direction. Phosphoramidite polynucleotide synthesis allows for the controlled addition of one nucleotide to a growing nucleic acid chain per synthesis cycle. In some instances, each synthesis cycle comprises a coupling step.

Phosphoramidite coupling involves the formation of a phosphite triester linkage between an activated nucleoside phosphoramidite and a nucleoside bound to the substrate, for example, via a linker. In some instances, the nucleoside phosphoramidite is provided to the device activated. In some instances, the nucleoside phosphoramidite is provided to the device with an activator. In some instances, nucleoside phosphoramidites are provided to the device in a 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100-fold excess or more over the substrate-bound nucleosides. In some instances, the addition of nucleoside phosphoramidite is performed in an anhydrous environment, for example, in anhydrous acetonitrile. Following addition of a nucleoside phosphoramidite, the device is optionally washed. In some instances, the coupling step is repeated one or more additional times, optionally with a wash step between nucleoside phosphoramidite additions to the substrate. In some instances, a polynucleotide synthesis method used herein comprises 1, 2, 3 or more sequential coupling steps. Prior to coupling, in many cases, the nucleoside bound to the device is de-protected by removal of a protecting group, where the protecting group functions to prevent polymerization. A common protecting group is 4,4'-dimethoxytrityl (DMT).

[00129] Following coupling, phosphoramidite polynucleotide synthesis methods optionally comprise a capping step. In a capping step, the growing polynucleotide is treated with a capping agent. A capping step is useful to block unreacted substrate-bound 5'-OH groups after coupling from further chain elongation, preventing the formation of polynucleotides with internal base deletions. Further, phosphoramidites activated with 1H-tetrazole may react, to a small extent, with the O6 position of guanosine. Without being bound by theory, upon oxidation with I₂/water, this side product, possibly via O6-N7 migration, may undergo depurination. The apurinic sites may end up being cleaved in the course of the final deprotection of the polynucleotide thus reducing the yield of the full-length product. The O6 modifications may be removed by treatment with the capping reagent prior to oxidation with I₂/water. In some instances, inclusion of a capping step during polynucleotide synthesis decreases the error rate as compared to synthesis without capping. As an example, the capping step comprises treating the substrate-bound polynucleotide with a mixture of acetic anhydride and 1-methylimidazole. Following a capping step, the device is optionally washed.

[00130] In some instances, following addition of a nucleoside phosphoramidite, and optionally after capping and one or more wash steps, the device bound growing nucleic acid is oxidized. The oxidation step comprises the phosphite triester is oxidized into a tetracoordinated phosphate triester, a protected precursor of the naturally occurring phosphate diester internucleoside linkage. In some instances, oxidation of the growing polynucleotide is achieved by treatment with iodine and water, optionally in the presence of a weak base (*e.g.*, pyridine, lutidine, collidine). Oxidation may be carried out under anhydrous conditions using, *e.g.* tert-Butyl hydroperoxide or (1S)-(+)-(10-camphorsulfonyl)-oxaziridine (CSO). In some methods, a capping step is performed following oxidation. A second capping step allows for device drying, as residual water from oxidation that may persist can inhibit subsequent coupling. Following oxidation, the device and growing polynucleotide is optionally washed. In some instances, the step of oxidation is substituted with a sulfurization step to obtain polynucleotide phosphorothioates, wherein any capping steps can be performed after the sulfurization. Many reagents are capable of the efficient sulfur transfer, including but not limited to 3-(Dimethylaminomethylidene)amino)-3H-1,2,4-dithiazole-3-thione, DDTT, 3H-1,2-benzodithiol-3-one 1,1-dioxide, also known as Beaucage reagent, and N,N,N',N'-Tetraethylthiuram disulfide (TETD).

[00131] In order for a subsequent cycle of nucleoside incorporation to occur through coupling, the protected 5' end of the device bound growing polynucleotide is removed so that the primary hydroxyl group is reactive with a next nucleoside phosphoramidite. In some instances, the

protecting group is DMT and deblocking occurs with trichloroacetic acid in dichloromethane. Conducting detritylation for an extended time or with stronger than recommended solutions of acids may lead to increased depurination of solid support-bound polynucleotide and thus reduces the yield of the desired full-length product. Methods and compositions of the disclosure described herein provide for controlled deblocking conditions limiting undesired depurination reactions. In some instances, the device bound polynucleotide is washed after deblocking. In some instances, efficient washing after deblocking contributes to synthesized polynucleotides having a low error rate.

[00132] Methods for the synthesis of polynucleotides typically involve an iterating sequence of the following steps: application of a protected monomer to an actively functionalized surface (*e.g.*, locus) to link with either the activated surface, a linker or with a previously deprotected monomer; deprotection of the applied monomer so that it is reactive with a subsequently applied protected monomer; and application of another protected monomer for linking. One or more intermediate steps include oxidation or sulfurization. In some instances, one or more wash steps precede or follow one or all of the steps.

[00133] Methods for phosphoramidite-based polynucleotide synthesis comprise a series of chemical steps. In some instances, one or more steps of a synthesis method involve reagent cycling, where one or more steps of the method comprise application to the device of a reagent useful for the step. For example, reagents are cycled by a series of liquid deposition and vacuum drying steps. For substrates comprising three-dimensional features such as wells, microwells, channels and the like, reagents are optionally passed through one or more regions of the device via the wells and/or channels.

[00134] Methods and systems described herein relate to polynucleotide synthesis devices for the synthesis of polynucleotides. The synthesis may be in parallel. For example, at least or about at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 10000, 50000, 75000, 100000 or more polynucleotides can be synthesized in parallel. The total number polynucleotides that may be synthesized in parallel may be from 2-100000, 3-50000, 4-10000, 5-1000, 6-900, 7-850, 8-800, 9-750, 10-700, 11-650, 12-600, 13-550, 14-500, 15-450, 16-400, 17-350, 18-300, 19-250, 20-200, 21-150, 22-100, 23-50, 24-45, 25-40, 30-35. Those of skill in the art appreciate that the total number of polynucleotides synthesized in parallel may fall within any range bound by any of these values, for example 25-100. The total number of polynucleotides synthesized in parallel may fall within any range defined by any of the values serving as endpoints of the range. Total molar mass of polynucleotides synthesized

within the device or the molar mass of each of the polynucleotides may be at least or at least about 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 25000, 50000, 75000, 100000 picomoles, or more. The length of each of the polynucleotides or average length of the polynucleotides within the device may be at least or about at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500 nucleotides, or more. The length of each of the polynucleotides or average length of the polynucleotides within the device may be at most or about at most 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 nucleotides, or less. The length of each of the polynucleotides or average length of the polynucleotides within the device may fall from 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35, 19-25. Those of skill in the art appreciate that the length of each of the polynucleotides or average length of the polynucleotides within the device may fall within any range bound by any of these values, for example 100-300. The length of each of the polynucleotides or average length of the polynucleotides within the device may fall within any range defined by any of the values serving as endpoints of the range.

[00135] Methods for polynucleotide synthesis on a surface provided herein allow for synthesis at a fast rate. As an example, at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200 nucleotides per hour, or more are synthesized. Nucleotides include adenine, guanine, thymine, cytosine, uridine building blocks, or analogs/modified versions thereof. In some instances, libraries of polynucleotides are synthesized in parallel on substrate. For example, a device comprising about or at least about 100; 1,000; 10,000; 30,000; 75,000; 100,000; 1,000,000; 2,000,000; 3,000,000; 4,000,000; or 5,000,000 resolved loci is able to support the synthesis of at least the same number of distinct polynucleotides, wherein polynucleotide encoding a distinct sequence is synthesized on a resolved locus. In some instances, a library of polynucleotides is synthesized on a device with low error rates described herein in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less. In some instances, larger nucleic acids assembled from a polynucleotide library synthesized with low error rate using the substrates and methods described herein are prepared in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less.

[00136] In some instances, methods described herein provide for generation of a library of nucleic acids comprising variant nucleic acids differing at a plurality of codon sites. In some instances, a nucleic acid may have 1 site, 2 sites, 3 sites, 4 sites, 5 sites, 6 sites, 7 sites, 8 sites, 9

sites, 10 sites, 11 sites, 12 sites, 13 sites, 14 sites, 15 sites, 16 sites, 17 sites, 18 sites, 19 sites, 20 sites, 30 sites, 40 sites, 50 sites, or more of variant codon sites.

[00137] In some instances, the one or more sites of variant codon sites may be adjacent. In some instances, the one or more sites of variant codon sites may not be adjacent and separated by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more codons.

[00138] In some instances, a nucleic acid may comprise multiple sites of variant codon sites, wherein all the variant codon sites are adjacent to one another, forming a stretch of variant codon sites. In some instances, a nucleic acid may comprise multiple sites of variant codon sites, wherein none the variant codon sites are adjacent to one another. In some instances, a nucleic acid may comprise multiple sites of variant codon sites, wherein some the variant codon sites are adjacent to one another, forming a stretch of variant codon sites, and some of the variant codon sites are not adjacent to one another.

[00139] Referring to the Figures, **FIG. 7** illustrates an exemplary process workflow for synthesis of nucleic acids (e.g., genes) from shorter nucleic acids. The workflow is divided generally into phases: (1) de novo synthesis of a single stranded nucleic acid library, (2) joining nucleic acids to form larger fragments, (3) error correction, (4) quality control, and (5) shipment. Prior to de novo synthesis, an intended nucleic acid sequence or group of nucleic acid sequences is preselected. For example, a group of genes is preselected for generation.

[00140] Once large nucleic acids for generation are selected, a predetermined library of nucleic acids is designed for de novo synthesis. Various suitable methods are known for generating high density polynucleotide arrays. In the workflow example, a device surface layer is provided. In the example, chemistry of the surface is altered in order to improve the polynucleotide synthesis process. Areas of low surface energy are generated to repel liquid while areas of high surface energy are generated to attract liquids. The surface itself may be in the form of a planar surface or contain variations in shape, such as protrusions or microwells which increase surface area. In the workflow example, high surface energy molecules selected serve a dual function of supporting DNA chemistry, as disclosed in International Patent Application Publication WO/2015/021080, which is herein incorporated by reference in its entirety.

[00141] In situ preparation of polynucleotide arrays is generated on a solid support and utilizes single nucleotide extension process to extend multiple oligomers in parallel. A deposition device, such as a material deposition device, is designed to release reagents in a step wise fashion such that multiple polynucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence **702**. In some instances,

polynucleotides are cleaved from the surface at this stage. Cleavage includes gas cleavage, *e.g.*, with ammonia or methylamine.

[00142] The generated polynucleotide libraries are placed in a reaction chamber. In this exemplary workflow, the reaction chamber (also referred to as “nanoreactor”) is a silicon coated well, containing PCR reagents and lowered onto the polynucleotide library 703. Prior to or after the sealing 704 of the polynucleotides, a reagent is added to release the polynucleotides from the substrate. In the exemplary workflow, the polynucleotides are released subsequent to sealing of the nanoreactor 705. Once released, fragments of single stranded polynucleotides hybridize in order to span an entire long range sequence of DNA. Partial hybridization 705 is possible because each synthesized polynucleotide is designed to have a small portion overlapping with at least one other polynucleotide in the pool.

[00143] After hybridization, a PCA reaction is commenced. During the polymerase cycles, the polynucleotides anneal to complementary fragments and gaps are filled in by a polymerase. Each cycle increases the length of various fragments randomly depending on which polynucleotides find each other. Complementarity amongst the fragments allows for forming a complete large span of double stranded DNA 706.

[00144] After PCA is complete, the nanoreactor is separated from the device 707 and positioned for interaction with a device having primers for PCR 708. After sealing, the nanoreactor is subject to PCR 709 and the larger nucleic acids are amplified. After PCR 710, the nanochamber is opened 711, error correction reagents are added 712, the chamber is sealed 713 and an error correction reaction occurs to remove mismatched base pairs and/or strands with poor complementarity from the double stranded PCR amplification products 714. The nanoreactor is opened and separated 715. Error corrected product is next subject to additional processing steps, such as PCR and molecular bar coding, and then packaged 722 for shipment 723.

[00145] In some instances, quality control measures are taken. After error correction, quality control steps include for example interaction with a wafer having sequencing primers for amplification of the error corrected product 716, sealing the wafer to a chamber containing error corrected amplification product 717, and performing an additional round of amplification 718. The nanoreactor is opened 719 and the products are pooled 720 and sequenced 721. After an acceptable quality control determination is made, the packaged product 722 is approved for shipment 723.

[00146] In some instances, a nucleic acid generate by a workflow such as that in FIG. 7 is subject to mutagenesis using overlapping primers disclosed herein. In some instances, a library

of primers are generated by in situ preparation on a solid support and utilize single nucleotide extension process to extend multiple oligomers in parallel. A deposition device, such as a material deposition device, is designed to release reagents in a step wise fashion such that multiple polynucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence **702**.

[00147] Computer systems

[00148] Any of the systems described herein, may be operably linked to a computer and may be automated through a computer either locally or remotely. In various instances, the methods and systems of the disclosure may further comprise software programs on computer systems and use thereof. Accordingly, computerized control for the synchronization of the dispense/vacuum/refill functions such as orchestrating and synchronizing the material deposition device movement, dispense action and vacuum actuation are within the bounds of the disclosure. The computer systems may be programmed to interface between the user specified base sequence and the position of a material deposition device to deliver the correct reagents to specified regions of the substrate.

[00149] The computer system **800** illustrated in **FIG. 8** may be understood as a logical apparatus that can read instructions from media **811** and/or a network port **805**, which can optionally be connected to server **809** having fixed media **812**. The system, such as shown in **FIG. 8** can include a CPU **801**, disk drives **803**, optional input devices such as keyboard **815** and/or mouse **816** and optional monitor **807**. Data communication can be achieved through the indicated communication medium to a server at a local or a remote location. The communication medium can include any means of transmitting and/or receiving data. For example, the communication medium can be a network connection, a wireless connection or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present disclosure can be transmitted over such networks or connections for reception and/or review by a party **822** as illustrated in **FIG. 8**.

[00150] **FIG. 14** is a block diagram illustrating a first example architecture of a computer system **1400** that can be used in connection with example instances of the present disclosure. As depicted in **FIG. 14**, the example computer system can include a processor **1402** for processing instructions. Non-limiting examples of processors include: Intel Xeon™ processor, AMD Opteron™ processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0™ processor, ARM Cortex-A8 Samsung S5PC100™ processor, ARM Cortex-A8 Apple A4™ processor, Marvell PXA 930™ processor, or a functionally-equivalent processor. Multiple threads of execution can be used for parallel processing. In some instances, multiple processors or processors with

multiple cores can also be used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[00151] As illustrated in **FIG. 9**, a high speed cache **904** can be connected to, or incorporated in, the processor **902** to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor **902**. The processor **902** is connected to a north bridge **906** by a processor bus **908**. The north bridge **906** is connected to random access memory (RAM) **910** by a memory bus **912** and manages access to the RAM **910** by the processor **902**. The north bridge **906** is also connected to a south bridge **914** by a chipset bus **916**. The south bridge **914** is, in turn, connected to a peripheral bus **918**. The peripheral bus can be, for example, PCI, PCI-X, PCI Express, or other peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data transfer between the processor, RAM, and peripheral components on the peripheral bus **918**. In some alternative architectures, the functionality of the north bridge can be incorporated into the processor instead of using a separate north bridge chip. In some instances, system **900** can include an accelerator card **922** attached to the peripheral bus **918**. The accelerator can include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator can be used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

[00152] Software and data are stored in external storage **924** and can be loaded into RAM **910** and/or cache **904** for use by the processor. The system **900** includes an operating system for managing system resources; non-limiting examples of operating systems include: Linux, WindowsTM, MACOSTM, BlackBerry OSTM, iOSTM, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example instances of the present disclosure. In this example, system **900** also includes network interface cards (NICs) **920** and **921** connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that can be used for distributed parallel processing.

[00153] **FIG. 10** is a diagram showing a network **1000** with a plurality of computer systems **1002a**, and **1002b**, a plurality of cell phones and personal data assistants **1002c**, and Network Attached Storage (NAS) **1004a**, and **1004b**. In example instances, systems **1002a**, **1002b**, and **1002c** can manage data storage and optimize data access for data stored in Network Attached Storage (NAS) **1004a** and **1004b**. A mathematical model can be used for the data and be

evaluated using distributed parallel processing across computer systems **1002a**, and **1002b**, and cell phone and personal data assistant systems **1002c**. Computer systems **1002a**, and **1002b**, and cell phone and personal data assistant systems **1002c** can also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) **1004a** and **1004b**. **FIG. 10** illustrates an example only, and a wide variety of other computer architectures and systems can be used in conjunction with the various instances of the present disclosure. For example, a blade server can be used to provide parallel processing. Processor blades can be connected through a back plane to provide parallel processing. Storage can also be connected to the back plane or as Network Attached Storage (NAS) through a separate network interface. In some example instances, processors can maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In other instances, some or all of the processors can use a shared virtual address memory space.

[00154] **FIG. 11** is a block diagram of a multiprocessor computer system **1100** using a shared virtual address memory space in accordance with an example instance. The system includes a plurality of processors **1102a-f** that can access a shared memory subsystem **1104**. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) **1106a-f** in the memory subsystem **1104**. Each MAP **1106a-f** can comprise a memory **1108a-f** and one or more field programmable gate arrays (FPGAs) **1110a-f**. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs **1110a-f** for processing in close coordination with a respective processor. For example, the MAPs can be used to evaluate algebraic expressions regarding the data model and to perform adaptive data restructuring in example instances. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP can use Direct Memory Access (DMA) to access an associated memory **1108a-f**, allowing it to execute tasks independently of, and asynchronously from the respective microprocessor **1102a-f**. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[00155] The above computer architectures and systems are examples only, and a wide variety of other computer, cell phone, and personal data assistant architectures and systems can be used in connection with example instances, including systems using any combination of general processors, co-processors, FPGAs and other programmable logic devices, system on chips (SOCs), application specific integrated circuits (ASICs), and other processing and logic elements. In some instances, all or part of the computer system can be implemented in software

or hardware. Any variety of data storage media can be used in connection with example instances, including random access memory, hard drives, flash memory, tape drives, disk arrays, Network Attached Storage (NAS) and other local or distributed data storage devices and systems.

[00156] In example instances, the computer system can be implemented using software modules executing on any of the above or other computer architectures and systems. In other instances, the functions of the system can be implemented partially or completely in firmware, programmable logic devices such as field programmable gate arrays (FPGAs) as referenced in FIG. 9, system on chips (SOCs), application specific integrated circuits (ASICs), or other processing and logic elements. For example, the Set Processor and Optimizer can be implemented with hardware acceleration through the use of a hardware accelerator card, such as accelerator card 922 illustrated in FIG. 9.

[00157] The following examples are set forth to illustrate more clearly the principle and practice of embodiments disclosed herein to those skilled in the art and are not to be construed as limiting the scope of any claimed embodiments. Unless otherwise stated, all parts and percentages are on a weight basis.

EXAMPLES

[00158] The following examples are given for the purpose of illustrating various embodiments of the disclosure and are not meant to limit the present disclosure in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the disclosure. Changes therein and other uses which are encompassed within the spirit of the disclosure as defined by the scope of the claims will occur to those skilled in the art.

[00159] Example 1: Functionalization of a device surface

[00160] A device was functionalized to support the attachment and synthesis of a library of polynucleotides. The device surface was first wet cleaned using a piranha solution comprising 90% H₂SO₄ and 10% H₂O₂ for 20 minutes. The device was rinsed in several beakers with DI water, held under a DI water gooseneck faucet for 5 min, and dried with N₂. The device was subsequently soaked in NH₄OH (1:100; 3 mL:300 mL) for 5 min, rinsed with DI water using a handgun, soaked in three successive beakers with DI water for 1 min each, and then rinsed again with DI water using the handgun. The device was then plasma cleaned by exposing the device surface to O₂. A SAMCO PC-300 instrument was used to plasma etch O₂ at 250 watts for 1 min in downstream mode.

[00161] The cleaned device surface was actively functionalized with a solution comprising N-(3-triethoxysilylpropyl)-4-hydroxybutyramide using a YES-1224P vapor deposition oven system with the following parameters: 0.5 to 1 torr, 60 min, 70 °C, 135 °C vaporizer. The device surface was resist coated using a Brewer Science 200X spin coater. SPR™ 3612 photoresist was spin coated on the device at 2500 rpm for 40 sec. The device was pre-baked for 30 min at 90 °C on a Brewer hot plate. The device was subjected to photolithography using a Karl Suss MA6 mask aligner instrument. The device was exposed for 2.2 sec and developed for 1 min in MSF 26A. Remaining developer was rinsed with the handgun and the device soaked in water for 5 min. The device was baked for 30 min at 100 °C in the oven, followed by visual inspection for lithography defects using a Nikon L200. A descum process was used to remove residual resist using the SAMCO PC-300 instrument to O₂ plasma etch at 250 watts for 1 min.

[00162] The device surface was passively functionalized with a 100 µL solution of perfluorooctyltrichlorosilane mixed with 10 µL light mineral oil. The device was placed in a chamber, pumped for 10 min, and then the valve was closed to the pump and left to stand for 10 min. The chamber was vented to air. The device was resist stripped by performing two soaks for 5 min in 500 mL NMP at 70 °C with ultrasonication at maximum power (9 on Crest system). The device was then soaked for 5 min in 500 mL isopropanol at room temperature with ultrasonication at maximum power. The device was dipped in 300 mL of 200 proof ethanol and blown dry with N₂. The functionalized surface was activated to serve as a support for polynucleotide synthesis.

[00163] Example 2: Synthesis of a 50-mer sequence on an oligonucleotide synthesis device

[00164] A two dimensional oligonucleotide synthesis device was assembled into a flowcell, which was connected to a flowcell (Applied Biosystems (ABI394 DNA Synthesizer"). The two-dimensional oligonucleotide synthesis device was uniformly functionalized with N-(3-TRIETHOXSILYLPROPYL)-4-HYDROXYBUTYRAMIDE (Gelest) was used to synthesize an exemplary polynucleotide of 50 bp ("50-mer polynucleotide") using polynucleotide synthesis methods described herein.

[00165] The sequence of the 50-mer was as described in SEQ ID NO.: 48.
 5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTT
 TTTT3' (SEQ ID NO.: 48), where # denotes Thymidine-succinyl hexamide CED
 phosphoramidite (CLP-2244 from ChemGenes), which is a cleavable linker enabling the release of oligos from the surface during deprotection.

[00166] The synthesis was done using standard DNA synthesis chemistry (coupling, capping, oxidation, and deblocking) according to the protocol in **Table 4** and an ABI synthesizer.

Table 4: Synthesis protocols

Table 4		
General DNA Synthesis Process Name	Process Step	Time (sec)
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N2 System Flush	4
	Acetonitrile System Flush	4
DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	6
	Activator + Phosphoramidite to Flowcell	6
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Incubate for 25sec	25
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	5
	Activator + Phosphoramidite to Flowcell	18
	Incubate for 25sec	25
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
CAPPING (CapA+B, 1:1, Flow)	CapA+B to Flowcell	15
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
OXIDATION (Oxidizer Flow)	Oxidizer to Flowcell	18

Table 4		
General DNA Synthesis Process Name	Process Step	Time (sec)
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N2 System Flush	4
	Acetonitrile System Flush	4
DEBLOCKING (Deblock Flow)	Deblock to Flowcell	36
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	18
	N2 System Flush	4.13
	Acetonitrile System Flush	4.13
	Acetonitrile to Flowcell	15

[00167] The phosphoramidite/activator combination was delivered similar to the delivery of bulk reagents through the flowcell. No drying steps were performed as the environment stays "wet" with reagent the entire time.

[00168] The flow restrictor was removed from the ABI 394 synthesizer to enable faster flow. Without flow restrictor, flow rates for amidites (0.1M in ACN), Activator, (0.25M Benzoylthiotetrazole ("BTT"; 30-3070-xx from GlenResearch) in ACN), and Ox (0.02M I2 in 20% pyridine, 10% water, and 70% THF) were roughly ~100uL/sec, for acetonitrile ("ACN") and capping reagents (1:1 mix of CapA and CapB, wherein CapA is acetic anhydride in THF/Pyridine and CapB is 16% 1-methylimidazole in THF), roughly ~200uL/sec, and for Deblock (3% dichloroacetic acid in toluene), roughly ~300uL/sec (compared to ~50uL/sec for all reagents with flow restrictor). The time to completely push out Oxidizer was observed, the timing for chemical flow times was adjusted accordingly and an extra ACN wash was introduced between different chemicals. After polynucleotide synthesis, the chip was deprotected in gaseous ammonia overnight at 75 psi. Five drops of water were applied to the surface to recover polynucleotides. The recovered polynucleotides were then analyzed on a BioAnalyzer small RNA chip.

[00169] Example 3: Synthesis of a 100-mer sequence on an oligonucleotide synthesis device

[00170] The same process as described in Example 2 for the synthesis of the 50-mer sequence was used for the synthesis of a 100-mer polynucleotide ("100-mer polynucleotide"; 5' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCA TGCTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTTTT3', where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes); SEQ ID NO.: 49) on two different silicon chips, the first one uniformly functionalized with N-(3-TRIETHOXYSYLYLPROPYL)-4-HYDROXYBUTYRAMIDE and the second one functionalized with 5/95 mix of 11-acetoxyundecyltriethoxysilane and n-decyltriethoxysilane, and the polynucleotides extracted from the surface were analyzed on a BioAnalyzer instrument.

[00171] All ten samples from the two chips were further PCR amplified using a forward (5'ATGCGGGGTTCTCATCATC3'; SEQ ID NO.: 50) and a reverse (5'CGGGATCCTTATCGTCATCG3'; SEQ ID NO.: 51) primer in a 50uL PCR mix (25uL NEB Q5 mastermix, 2.5uL 10uM Forward primer, 2.5uL 10uM Reverse primer, 1uL polynucleotide extracted from the surface, and water up to 50uL) using the following thermalcycling program:

98 °C, 30 sec

98 °C, 10 sec; 63 °C, 10 sec; 72 °C, 10 sec; repeat 12 cycles

72 °C, 2min

[00172] The PCR products were also run on a BioAnalyzer, demonstrating sharp peaks at the 100-mer position. Next, the PCR amplified samples were cloned, and Sanger sequenced. **Table 5** summarizes the results from the Sanger sequencing for samples taken from spots 1-5 from chip 1 and for samples taken from spots 6-10 from chip 2.

Table 5: Sequencing results

Spot	Error rate	Cycle efficiency
1	1/763 bp	99.87%
2	1/824 bp	99.88%
3	1/780 bp	99.87%
4	1/429 bp	99.77%
5	1/1525 bp	99.93%
6	1/1615 bp	99.94%
7	1/531 bp	99.81%

Spot	Error rate	Cycle efficiency
8	1/1769 bp	99.94%
9	1/854 bp	99.88%
10	1/1451 bp	99.93%

[00173] Thus, the high quality and uniformity of the synthesized polynucleotides were repeated on two chips with different surface chemistries. Overall, 89% of the 100-mers that were sequenced were perfect sequences with no errors, corresponding to 233 out of 262.

[00174] Table 6 summarizes error characteristics for the sequences obtained from the polynucleotides samples from spots 1-10.

Table 6: Error characteristics

Sample ID/Spot no.	OSA_0 046/1	OSA_0 047/2	OSA_0 048/3	OSA_0 049/4	OSA_0 050/5	OSA_0 051/6	OSA_0 052/7	OSA_0 053/8	OSA_0 054/9	OSA_00 55/10
Total Sequences	32	32	32	32	32	32	32	32	32	32
Sequencing Quality	25 of 28	27 of 27	26 of 30	21 of 23	25 of 26	29 of 30	27 of 31	29 of 31	28 of 29	25 of 28
Oligo Quality	23 of 25	25 of 27	22 of 26	18 of 21	24 of 25	25 of 29	22 of 27	28 of 29	26 of 28	20 of 25
ROI Match Count	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI Mutation	2	2	1	3	1	0	2	1	2	1
ROI Multi Base Deletion	0	0	0	0	0	0	0	0	0	0
ROI Small Insertion	1	0	0	0	0	0	0	0	0	0
ROI Single Base Deletion	0	0	0	0	0	0	0	0	0	0
Large Deletion Count	0	0	1	0	0	1	1	0	0	0
Mutation: G>A	2	2	1	2	1	0	2	1	2	1
Mutation: T>C	0	0	0	1	0	0	0	0	0	0
ROI Error Count	3	2	2	3	1	1	3	1	2	1
ROI Error Rate	Err: ~1 in 834	Err: ~1 in 1350	Err: ~1 in 1282	Err: ~1 in 708	Err: ~1 in 2500	Err: ~1 in 2667	Err: ~1 in 876	Err: ~1 in 2900	Err: ~1 in 1400	Err: ~1 in 2349

ROI	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP Err:
Minus	Err: ~1	Err: ~1	Err: ~1	Err: ~1	Err: ~1	Err: ~1	Err: ~1	Err: ~1	Err: ~1	Err: ~1	~1 in
Primer	in 763	in 824	in 780	in 429	in 1525	in 1615	in 531	in 1769	in 854		1451
Error Rate											

[00175] Example 4: Design of G protein-coupled receptor binding domains based on conformational ligand interactions

[00176] G protein-coupled receptor (GPCR) binding domains were designed using interaction surfaces between conformational ligands that interact with GPCRs. Analysis of the interaction surfaces between chemokines and cytokines and the GPCRs indicated that the N-terminal peptide prior to the first conformational cysteine represents the activation peptide, and the core helical and beta-turn-beta topologies mediate interactions with the extracellular domain (ECD) of the GPCR.

[00177] An additional 254 GPCR ligands were designed based on cross-searching Uniprot and IUPHAR databases. The ligands represented 112 human, 71 rat, 4 pig, 1 sheep, and 1 cow derived interaction classes. The ligands were then collapsed to the following 101 cross-species ligand sequence annotations: ADM, ADM2, Agouti-related protein, Angiotensinogen, Annexin A1, Apelin, Apelin receptor early, Appetite regulating hormone, Beta-defensin 4A, C-C motif chemokine, C-X-C motif chemokine, Calcitonin, Calcitonin gene-related peptide, Cathepsin G, Cathepsin G (Fragment), Cholecystokinin, Complement C3, Complement C5, Complement C5 (Fragment), Corticoliberin, Cortistatin, Cytokine SCM-1 beta, Endothelin-2, Endothelin-3, Eotaxin, Fractalkine, Galanin peptides, Galanin-like peptide, Gastric inhibitory polypeptide, Gastrin, Gastrin-releasing peptide, Glucagon, Growth-regulated alpha protein, Heme-binding protein 1, Humanin, Insulin-like 3, Insulin-like peptide INSL5, Interleukin-8, Islet amyloid polypeptide, Kininogen-1, Lymphotoxin, Metastasis-suppressor KiSS-1, Neurokinin-B, Neuromedin-B, Neuromedin-S, Neuromedin-U, Neuropeptide B, Neuropeptide S, Neuropeptide W, Neurotensin/neuromedin N, Orexigenic neuropeptide QRFP, Orexin, Oxytocin-neurophysin 1, Pancreatic prohormone, Parathyroid hormone, Parathyroid hormone-related protein, Peptide YY, Pituitary adenylate cyclase-activating, Platelet basic protein, Platelet factor 4, Prepronociceptin, Pro-FMRFamide-related neuropeptide FF, Pro-FMRFamide-related neuropeptide VF, Pro-MCH, Pro-neuropeptide Y, Pro-opiomelanocortin, Pro-thyrotropin releasing hormone, Proenkephalin-A, Proenkephalin-B, Progonadoliberin-1, Progonadoliberin-2, Prokineticin-1, Prokineticin-2, Prolactin-releasing peptide, Promotilin, Protachykinin-1, Protein Wnt-2, Protein Wnt-3a, Protein Wnt-4, Protein Wnt-5a, Protein Wnt-7b, Prothrombin, Proto-oncogene Wnt-1, Protooncogene Wnt-3, Putative uncharacterized protein, RCG55748, Retinoic acid receptor, Secretin, Somatoliberin, Somatostatin, Stromal cell-derived factor, T-kininogen 2,

Tuberoinfundibular peptide of, Urocortin, Urocortin-2, Urocortin-3, Urotensin-2, Urotensin-2B, VEGF coregulated chemokine, VIP peptides, and Vasopressin-neurophysin 2-copeptin.

[00178] Structural analysis of the ligands was performed and indicated that a majority of them comprised the N-terminal activation peptide of about 11 amino acids. Motif variants were then created by trimming back the N-terminal activation peptide. As seen in **Table 7**, an exemplary set of variants were created based on the N-terminal activation peptide for stromal derived factor-1. The motif variants were also placed combinatorially at multiple positions in the CDR-H3. A total of 1016 motifs were extracted for placement in the CDR-H3. In addition, the motif variants were provided with variably boundary placement and with 5-20 substring variants that were also placed in the CDR-H3.

Table 7. Variant amino acid sequences for stromal derived factor-1

SEQ ID NO.	Variant	Amino Acid Sequence
53	1	ggggSDYKPVLSYSR
54	2	ggggDYKPVLSYSR
55	3	ggggYKPVLSYSR
56	4	ggggKPVLSYSR
57	5	ggggPVLSYSR

[00179] As seen in **Table 8**, an exemplary set of variants were created for interleukin-8 based on the following sequence:

MTSKLAVALLAAFLISAALCEGAVLPRSAKELRCQCIKTYSKPFHPKFIKELRVIESGPH
CANTEIIVKLSDBGRELCLDPKENWVQRVVEKFLKRAENS (SEQ ID NO: 52).

Table 8. Variant amino acid sequences for interleukin-8

SEQ ID NO.	Variant	Amino Acid Sequence
58	1	ggggSAALCEGAVLPRSA
59	2	ggggAALCEGAVLPRSA
60	3	ggggALCEGAVLPRSA
61	4	ggggLCEGAVLPRSA
62	5	ggggCEGAVLPRSA
63	6	ggggSAALCEGAVLPRSAKE
64	7	ggggAALCEGAVLPRSAKE
65	8	ggggALCEGAVLPRSAKE
66	9	ggggLCEGAVLPRSAKE
67	10	ggggCEGAVLPRSAKE
68	11	ggggSAALCEGAVLPRSAKELR
69	12	ggggAALCEGAVLPRSAKELR
70	13	ggggALCEGAVLPRSAKELR
71	14	ggggLCEGAVLPRSAKELR
72	15	ggggCEGAVLPRSAKELR

[00180] Example 5: Design of G protein-coupled receptor binding domains based on peptide ligand interactions

[00181] GPCR binding domains were designed based on interaction surfaces between peptide ligands that interact with class B GPCRs. About 66 different ligands were used and include the following ligand sequence annotations: Adrenomedullin, Amylin, Angiotensin, Angiotensin I, Angiotensin II, Angiotensin III, Apelin, Apstatin, Big Endothelin, Big Gastrin, Bradykinin, Caerulein, Calcitonin, Calcitonin Gene Related Peptide, CGRP, Cholecystokinin, Endothelin, Endothelin 1, Endothelin 2, Endothelin 3, GIP, GIPs, GLP, Galanin, Gastrin, Ghrelin, Glucagon, IAPP, Kisspeptin, Mca, Metasti, Neuromedin, Neuromedin N, Neuropeptide, Neuropeptide F, Neuropeptide Y, Neurotensin, Nociceptin, Orexin, Orexin A, Orphanin, Oxytocin, Oxytocin Galanin, PACAP, PACAPs, PAR (Protease Activated Receptor) Peptides, PAR-1 Agonist, Pramlintide, Scyliorhinin I, Secretin, Senktide, Somatostatin, Somatostatin 14, Somatostatin 28, Substance P, Urotensin II, VIP, VIPs, Vasopressin, Xenin, cinnamoyl, furoyl, gastrin, holecystokinin, α -Mating Factor Pheromone. It was observed that the peptides formed a stabilized interaction with the GPCR extracellular domain (ECD).

[00182] Motif variants were generated based on the interaction surface of the peptides with the ECD as well as with the N-terminal GPCR ligand interaction surface. This was done using structural modeling. Exemplary motif variants were created based on glucagon like peptide's interaction with its GPCR as seen in **Table 9**. The motif variant sequences were generated using the following sequence from glucagon like peptide:

HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG (SEQ ID NO: 73).

Table 9. Variant amino acid sequences for glucagon like peptide

SEQ ID NO.	Variant	Amino Acid Sequence
74	1	sgggsgsgsgsgsggHAEGTFTSDVSSYLEGQAAKEFIAWL
75	2	sgggsgsgsgsgsggAEGTFTSDVSSYLEGQAAKEFIAWL
76	3	sgggsgsgsgsgsggEGTFTSDVSSYLEGQAAKEFIAWL
77	4	sgggsgsgsgsgsggGTFTSDVSSYLEGQAAKEFIAWL
78	5	sgggsgsgsgsgsggTFTSDVSSYLEGQAAKEFIAWL
79	6	sgggsgsgsgsgsggFTSDVSSYLEGQAAKEFIAWL
80	7	sgggsgsgsgsgsggTSDVSSYLEGQAAKEFIAWL
81	8	sgggsgsgsgsgsggSDVSSYLEGQAAKEFIAWL
82	9	sgggsgsgsgsgsggDVSSYLEGQAAKEFIAWL

[00183] Example 6: Design of G protein-coupled receptor binding domains based on small molecule interactions

[00184] GPCR binding domains were designed based on interaction surfaces between small molecule ligands that interact with GPCRs. By analyzing multiple GPCR ligands, an amino acid library of Tyr, Pro, Phe, His, and Gly was designed as being able to recapitulate many of the structural contacts of these ligands. An exemplary motif variant that was generated based on these observations comprises the following sequence:

sgggg(F,G,H,P,Y)(F,G,H,P,Y)(F,G,H,P,Y)(F,G,H,P,Y)(F,G,H,P,Y)(F,G,H,P,Y)(F,G,H,P,Y)(F,G,H,P,Y) (SEQ ID NO: 83).

[00185] Example 7: Design of G protein-coupled binding domains based on extracellular domain interactions

[00186] GPCR binding domains were designed based on interaction surfaces on extracellular domains (ECDs) and extracellular loops (ECLs) of GPCRs. About 2,257 GPCRs from human (356), mouse (369), rat (259), cow (102), pig (60), primate, fish, fly, and over 200 other organisms were analyzed, and it was observed that ECDs provide multiple complementary contacts to other loops and helices of the GPCR at a length of 15 amino acids. Further analysis of the ECLs from the about 2,257 GPCRs and all solved structures of GPCRs demonstrated that the N-terminal ECD1 and ECL2 comprise longer extracellular sequences and provide GPCR extracellular contacts.

[00187] Motif variants were then generated based on these sequences. Exemplary variants based on the following sequence from retinoic acid induced protein 3 (GPRC5A) were generated: EYIVLTMNRTNVNVFSELSAPRRNED (SEQ ID NO: 84). *See Table 10.*

Table 10. Amino acid sequences

SEQ ID NO.	Variant	Amino Acid Sequence
85	1	YIVLTMNRTNVNVFSELSAPRRNE
86	2	IVLTMNRTNVNVFSELSAPRRN
87	3	VLTMNRTNVNVFSELSAPRR

[00188] Example 8: Design of antibody scaffolds

[00189] To generate scaffolds, structural analysis, repertoire sequencing analysis of the heavy chain, and specific analysis of heterodimer high-throughput sequencing datasets were performed. Each heavy chain was associated with each light chain scaffold. Each heavy chain scaffold was assigned 5 different long CDR-H3 loop options. Each light chain scaffold was assigned 5 different L3 scaffolds. The heavy chain CDR-H3 stems were chosen from the frequently observed long H3 loop stems (10 amino acids on the N-terminus and the C-terminus) found both across individuals and across V-gene segments. The light chain scaffold L3s were chosen from heterodimers comprising long H3s. Direct heterodimers based on information from the Protein

Data Bank (PDB) and deep sequencing datasets were used in which CDR H1, H2, L1, L2, L3, and CDR-H3 stems were fixed. The various scaffolds were then formatted for display on phage to assess for expression.

[00190] *Structural Analysis*

[00191] About 2,017 antibody structures were analyzed from which 22 structures with long CDR-H3s of at least 25 amino acids in length were observed. The heavy chains included the following: IGHV1-69, IGHV3-30, IGHV4-49, and IGHV3-21. The light chains identified included the following: IGLV3-21, IGKV3-11, IGKV2-28, IGKV1-5, IGLV1-51, IGLV1-44, and IGKV1-13. In the analysis, four heterodimer combinations were observed multiple times including: IGHV4-59/61-IGLV3-21, IGHV3-21-IGKV2-28, IGHV1-69-IGKV3-11, and IGHV1-69-IGKV1-5. An analysis of sequences and structures identified intra-CDR-H3 disulfide bonds in a few structures with packing of bulky side chains such as tyrosine in the stem providing support for long H3 stability. Secondary structures including beta-turn-beta sheets and a “hammerhead” subdomain were also observed.

[00192] *Repertoire Analysis*

[00193] A repertoire analysis was performed on 1,083,875 IgM+/CD27-naïve B cell receptor (BCR) sequences and 1,433,011 CD27+ sequences obtained by unbiased 5'RACE from 12 healthy controls. The 12 healthy controls comprised equal numbers of male and female and were made up of 4 Caucasian, 4 Asian, and 4 Hispanic individuals. The repertoire analysis demonstrated that less than 1% of the human repertoire comprises BCRs with CDR-H3s longer than 21 amino acids. A V-gene bias was observed in the long CDR3 subrepertoire, with IGHV1-69, IGHV4-34, IGHV1-18, and IGHV1-8 showing preferential enrichment in BCRs with long H3 loops. A bias against long loops was observed for IGHV3-23, IGHV4-59/61, IGHV5-51, IGHV3-48, IGHV3-53/66, IGHV3-15, IGHV3-74, IGHV3-73, IGHV3-72, and IGHV2-70. The IGHV4-34 scaffold was demonstrated to be autoreactive and had a short half-life.

[00194] Viable N-terminal and C-terminal CDR-H3 scaffold variation for long loops were also designed based on the 5'RACE reference repertoire. About 81,065 CDR-H3s of amino acid length 22 amino acids or greater were observed. By comparing across V-gene scaffolds, scaffold-specific H3 stem variation was avoided as to allow the scaffold diversity to be cloned into multiple scaffold references.

[00195] *Heterodimer Analysis*

[00196] Heterodimer analysis was performed on scaffolds having sequences as seen in **FIGS. 12A-12C**. Variant sequences and lengths of the scaffolds were assayed.

[00197] *Structural Analysis*

[00198] Structural analysis was performed using GPCR scaffolds of variant sequences and lengths were assayed. See FIG. 13.

[00199] **Example 9: Generation of GPCR antibody libraries**

[00200] Based on GPCR-ligand interaction surfaces and scaffold arrangements, libraries were designed and de novo synthesized. See Examples 4-8. Referring to FIG. 5, 10 variant sequences were designed for the variable domain, heavy chain 503, 237 variant sequences were designed for the heavy chain complementarity determining region 3 507, and 44 variant sequences were designed for the variable domain, light chain 513. The fragments were synthesized as three fragments as seen in FIG. 6 following similar methods as described in Examples 1-3.

[00201] Following de novo synthesis, 10 variant sequences were generated for the variable domain, heavy chain 602, 236 variant sequences were generated for the heavy chain complementarity determining region 3 604, and 43 variant sequences were designed for a region comprising the variable domain 606, light chain and CDR-L3 and of which 9 variants for variable domain, light chain were designed. This resulted in a library with about 10^5 diversity ($10 \times 236 \times 43$). This was confirmed using next generation sequencing (NGS) with 16 million reads. As seen in FIG. 14, the normalized sequencing reads for each of the 10 variants for the variable domain, heavy chain was about 1. As seen in FIG. 15, the normalized sequencing reads for each of the 43 variants for the variable domain, light chain was about 1. As seen in FIG. 16, the normalized sequencing reads for 236 variant sequences for the heavy chain complementarity determining region 3 were about 1.

[00202] The various light and heavy chains were then tested for expression and protein folding. Referring to FIGS. 17A-17D, the 10 variant sequences for variable domain, heavy chain included the following: IGHV1-18, IGHV1-69, IGHV1-8 IGHV3-21, IGHV3-23, IGHV3-30/33m, IGHV3-28, IGHV3-74, IGHV4-39, and IGHV4-59/61. Of the 10 variant sequences, IGHV1-18, IGHV1-69, and IGHV3-30/33m exhibited improved characteristics such as improved thermostability. Referring to FIGS. 18A-18F, 9 variant sequences for variable domain, light chain included the following: IGKV1-39, IGKV1-9, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV4-1, IGLV1-51, and IGLV2-14. Of the 9 variant sequences, IGKV1-39, IGKV3-15, IGLV1-51, and IGLV2-14 exhibited improved characteristics such as improved thermostability.

[00203] **Example 10: Expression of GPCR antibody libraries in HEK293 cells**

[00204] Following generation of GPCR antibody libraries as in Example 13, about 47 GPCRs were selected for screening. GPCR constructs about 1.8 kb to about 4.5 kb in size were designed in a pCDNA3.1 vector. The GPCR constructs were then synthesized following similar methods

as described in Examples 2-4 including hierarchical assembly. Of the 47 GPCR constructs, 46 GPCR constructs were synthesized.

[00205] The synthesized GPCR constructs were transfected in HEK293 and assayed for expression using immunofluorescence. Referring to **FIGS. 19A-19C**, HEK293 cells were transfected with the GPCR constructs comprising an N-terminally hemagglutinin (HA)-tagged human Y₁ receptor. Following 24-48 hours of transfection, cells were washed with phosphate buffered saline (PBS) and fixed with 4% paraformaldehyde. Cells were stained using fluorescent primary antibody directed towards the HA tag or secondary antibodies comprising a fluorophore and DAPI to visualize the nuclei in blue. Referring to **FIGS. 19A-19C**, human Y₁ receptor was visualized on the cell surface in non-permeabilized cells and on the cell surface and intracellularly in permeabilized cells.

[00206] GPCR constructs were also visualized by designing GPCR constructs comprising auto-fluorescent proteins. Referring to **FIGS. 20A-20C**, human Y₁ receptor comprised EYFP fused to its C-terminus, and human Y₅ receptor comprised ECFP fused to its C-terminus. HEK293 cells were transfected with human Y₁ receptor or co-transfected with human Y₁ receptor and human Y₅ receptor. Following transfection cells were washed and fixed with 4% paraformaldehyde. Cells were stained with DAPI. Localization of human Y₁ receptor and human Y₅ receptor were visualized by fluorescence microscopy.

[00207] **Example 11: Design of immunoglobulin library**

[00208] An immunoglobulin scaffold library was designed for placement of GPCR binding domains and for improving stability for a range of GPCR binding domain encoding sequences. The immunoglobulin scaffold included a VH domain attached with a VL domain with a linker. Variant nucleic acid sequences were generated for the framework elements and CDR elements of the VH domain and VL domain. The structure of the design is shown in **FIG. 21A**. A full domain architecture is shown in **FIG. 21B**. Sequences for the leader, linker, and pIII are listed in **Table 11**.

Table 11. Nucleotide sequences

SEQ ID NO	Domain	Sequence
88	Leader	GCAGCCGCTGGCTTGCTGCTGCTGGCAGCTCAGCCGGCCATGGCC
89	Linker	GCTAGCGGTGGAGGCGGTTTCAGGCGGAGGTGGCTCTGGCGGTGGC GGATCGCATGCATCC
90	pIII	CGCGCGGCCGCTGGAAGCGGCTCCCACCATCACCATCACCAT

[00209] The VL domains that were designed include IGKV1-39, IGKV3-15, IGLV1-51, and IGLV2-14. Each of four VL domains were assembled with their respective invariant four framework elements (FW1, FW2, FW3, FW4) and variable 3 CDR (L1, L2, L3) elements. For IGKV1-39, there was 490 variants designed for L1, 420 variants designed for L2, and 824 variants designed for L3 resulting in a diversity of 1.7×10^8 ($490 * 420 * 824$). For IGKV3-15, there was 490 variants designed for L1, 265 variants designed for L2, and 907 variants designed for L3 resulting in a diversity of 1.2×10^8 ($490 * 265 * 907$). For IGLV1-51, there was 184 variants designed for L1, 151 variants designed for L2, and 824 variants designed for L3 resulting in a diversity of 2.3×10^7 ($184 * 151 * 824$). IGLV2-14, 967 variants designed for L1, 535 variants designed for L2, and 922 variants designed for L3 resulting in a diversity of 4.8×10^8 ($967 * 535 * 922$). **Table 12** lists the amino acid sequences and nucleotide sequences for the four framework elements (FW1, FW2, FW3, FW4) for IGLV1-51. **Table 13** lists the variable 3 CDR (L1, L2, L3) elements for IGLV1-51. Variant amino acid sequences and nucleotide sequences for the four framework elements (FW1, FW2, FW3, FW4) and the variable 3 CDR (L1, L2, L3) elements were also designed for IGKV1-39, IGKV3-15, and IGLV2-14.

Table 12. Sequences for IGLV1-51 framework elements

Element	SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Nucleotide Sequence
IGLV1-51				
FW1	91	QSVLTQPPSVSAAPGQKV TISC	92	CAGTCTGTGTTGACGCAGCCGCC TCAGTGTCTGCGGCCCCAGGACAG AAGGTCACCATCTCCTGC
FW2	93	WYQQLPGTAPKLLIY	94	TGGTATCAGCAGCTCCCAGGAACA GCCCCAAACTCCTCATTAT
FW3	95	GIPDRFSGSGSGTSATLGI TGLQTGDEADYY	96	GGGATTCTGACCGATTCTCTGGC TCCAAGTCTGGCACGTCAGCCACC CTGGGCATCACCGGACTCCAGACT GGGGACGAGGCCGATTATTAC
FW4	97	GGGTKLTVL	98	GGCGGAGGGACCAAGCTGACCGTC CTA

Table 13. Sequences for IGLV1-51 CDR elements

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Nucleotide Sequence
IGLV1-51-L1			
99	SGSSSNIGS NHVS	282	TCTGGAAGCAGCTCCAACATTGGGAGTAATCATGT ATCC
100	SGSSSNIGN NYLS	283	TCTGGAAGCAGCTCCAACATTGGGAATAATTATCT ATCC

101	SGSSSNIAN NYVS	284	TCTGGAAGCAGCTCCAACATTGCGAATAATTATGT ATCC
102	SGSSPNIGN NYVS	285	TCTGGAAGCAGCCCCAACATTGGGAATAATTATGT ATCG
103	SGSRSNIGS NYVS	286	TCTGGAAGCAGATCCAATATTGGGAGTAATTATGT TTCG
104	SGSSSNVG DNYVS	287	TCTGGAAGCAGCTCCAACGTTGGCGATAATTATGT TTCC
105	SGSSSNIGI QYVS	288	TCTGGAAGCAGCTCCAACATTGGGATTCAATATGT ATCC
106	SGSSSNVG NNFVS	289	TCTGGAAGCAGCTCCAATGTTGGTAACAATTTGTCTCC
107	SGSASNIGN NYVS	290	TCTGGAAGCGCCTCCAACATTGGGAATAATTATGT ATCC
108	SGSGSNIGN NDVS	291	TCTGGAAGCGGCTCCAATATTGGGAATAATGATGT GTCC
109	SGSISNIGN NYVS	292	TCTGGAAGCATCTCCAACATTGGTAATAATTATGT ATCC
110	SGSISNIGK NYVS	293	TCTGGAAGCATCTCCAACATTGGGAAAAATTATGT GTCG
111	SGSSSNIGH NYVS	294	TCTGGAAGCAGCTCCAACATTGGGCATAATTATGT ATCG
112	PGSSSNIGN NYVS	295	CCTGGAAGCAGCTCCAACATTGGGAATAATTATGT ATCC
113	SGSTSNIGI HYVS	296	TCTGGAAGCACCTCCAACATTGGAATTCATTATGT ATCC
114	SGSSSNIGS HYVS	297	TCTGGAAGCAGCTCCAACATTGGCAGTCATTATGT TTCC
115	SGSSSNIGN EYVS	298	TCCGGAAGCAGCTCCAACATTGGAAATGAATATGT ATCC
116	SGSTSNIGN NYIS	299	TCTGGAAGCACCTCCAACATTGGGAAATAATTATAT ATCG
117	SGSSSNIGN HFVS	300	TCTGGAAGCAGCTCCAATATTGGGAATCATTTTGT ATCG
118	SGSSSNIGN NYVA	301	TCTGGAAGCAGCTCCAACATTGGGAATAATTATGT GGCC
119	SGSSSNIGS YYVS	302	TCTGGAAGCAGCTCCAACATTGGAAGTTATTATGT ATCC
120	SGSGFNIGN NYVS	303	TCTGGAAGTGTTTCAACATTGGGAATAATTATGT CTCT
121	SGSTSNIGN NYVS	304	TCTGGAAGCACCTCCAACATTGGGAATAATTATGT GTCC
122	SGSSSDIGN NYVS	305	TCTGGAAGCAGCTCCGACATTGGCAATAATTATGT ATCC
123	SGSSSNIGN NVVS	306	TCTGGAAGCAGCTCCAACATTGGGAATAATGTTGT ATCC
124	SGSKSNIGK NYVS	307	TCTGGAAGCAAGTCTAACATTGGGAAAAATTATGT ATCC
125	SGSSTNIGN NYVS	308	TCTGGAAGCAGCACCAACATTGGGAATAATTATGT ATCC

126	SGSISNIGD NYVS	309	TCTGGAAGCATCTCCAACATTGGGGATAATTATGT ATCC
127	SGSSSNIGS KDVS	310	TCTGGAAGCAGCTCCAACATTGGGAGTAAGGATGT ATCA
128	SGSSSNIGS NDVS	311	TCTGGAAGCAGCTCCAACATTGAGAATAATTATGT ATCG
129	SGSSSNIGN HYVS	312	TCTGGAAGCAGCTCCAACATTGGGAATCATTATGT ATCC
130	SGSSSNIGK DFVS	313	TCTGGAAGCAGCTCCAACATTGGGAAGGATTTTGT CTCC
131	SGSTSNIGS NFVS	314	TCTGGCAGTACTTCCAACATCGGAAGTAATTTTGT TCC
132	SGSTSNIGH NYVS	315	TCTGGAAGCACCTCCAACATTGGGCATAATTATGT ATCC
133	SASSSNIGN NYVS	316	TCTGCAAGCAGCTCCAACATTGGGAATAATTATGT ATCC
134	SGSSSIGN NYVS	317	TCTGGAAGCAGCTCCAGCATTGGCAATAATTATGT ATCC
135	SGSSSTIGN NYVS	318	TCTGGAAGCAGCTCCACCATTGGGAATAATTATGT ATCC
136	SGSSSNIGS NYVS	319	TCTGGAAGCAGCTCCAACATTGAAAATAATTATGT ATCC
137	SGSSSNIGN QYVS	320	TCTGGAAGCAGCTCCAACATTGGGAATCAGTATGT ATCC
138	SGSSSNIGN NYVF	321	TCTGGAAGCAGCTCCAACATTGGGAATAATTATGT ATCC
139	SGSSSNIGR NYVS	322	TCTGGAAGCAGCTCCAACATTGGGAGGAATTATGT CTCC
140	SGGSSNIGN YYVS	323	TCTGGAGGCAGCTCCAACATTGGAAATTATTATGT ATCG
141	SGSSSNIGD NYVS	324	TCTGGAAGCAGCTCCAACATTGGAGATAATTATGT CTCC
142	SGGSSNIGI NYVS	325	TCTGGAGGCAGCTCCAACATTGGAATTAATTATGT ATCC
143	SGGSSNIGK NYVS	326	TCTGGAGGCAGCTCCAACATTGGGAAGAATTATGT ATCC
144	SGSSSNIGK RSVS	327	TCTGGAAGCAGCTCCAACATTGGGAAGAGATCTGT ATCG
145	SGSRSNIGN NYVS	328	TCTGGAAGCAGATCCAACATTGGGAATAACTATGT ATCC
146	SGSSSNIGN NLVS	329	TCGGGAAGCAGCTCCAACATTGGGAATAATCTTGT TTCC
147	SGSSSNIGI NYVS	330	TCTGGAAGCAGCTCCAACATTGGGATCAATTATGT ATCC
148	SGSSSNIGN NFVS	331	TCTGGAAGCAGCTCCAACATCGGGAATAATTTTGT ATCC
149	SGTSSNIGR NFVS	332	TCTGGAACCAGCTCCAACATTGGCAGAAATTTTGT ATCC
150	SGRRSNIG NNYVS	333	TCTGGAAGGAGGTCCAACATTGGAAATAATTATGT GTCC

151	SGGSFNIGN NYVS	334	TCTGGAGGCAGCTTCAATATTGGGAATAATTATGT ATCC
152	SGSTSNIGE NYVS	335	TCTGGAAGCACTTCCAACATTGGGGAGAATTATGT GTCC
153	SGSSSNIGS DYVS	336	TCTGGAAGCAGCTCCAATATTGGGAGTGATTATGT ATCC
154	SGTSSNIGS NYVS	337	TCTGGAACCAGCTCCAACATTGGGAGTAATTATGT ATCC
155	SGSSSNIGT NFVS	338	TCTGGAAGCAGCTCCAACATTGGGACTAATTTTGT ATCC
156	SGSSSNFG NNYVS	339	TCTGGAAGCAGCTCCAACTTTGGGAATAATTATGT ATCC
157	SGSTSNIGN NHVS	340	TCTGGAAGCACCTCCAACATTGGGAATAATCATGT ATCC
158	SGSSSNIGN DFVS	341	TCTGGAAGCAGCTCCAACATTGGGAATGATTTTGT ATCC
159	SGSSSDIGD NYVS	342	TCTGGAAGCAGCTCCGACATTGGCGATAATTATGT GTCC
160	SGSSSNIGK YYVS	343	TCTGGAAGCAGCTCCAACATTGGGAAATATTATGT ATCC
161	SGSSSNIGG NYVS	344	TCTGGAAGCAGCTCCAACATTGGCGGTAATTATGT ATCC
162	SGSSSNTG NNYVS	345	TCTGGAAGCAGCTCCAACACTGGGAATAATTATGT ATCC
163	SGSSSNVG NNYVS	346	TCTGGAAGCAGCTCCAACGTTGGGAATAATTATGT GTCT
164	SGSSSNIAN NFVS	347	TCTGGAAGCAGCTCCAACATTGCGAATAATTTTGT ATCC
165	SGSSSNIGN DYVS	348	TCTGGAAGCAGCTCCAACATTGGGAATGATTATGT ATCC
166	SGSTSNIE NYVS	349	TCTGGAAGCACCTCCAATATTGAGAATAATTATGT TTCC
167	SGGSSNIGN NDVS	350	TCTGGAGGCAGCTCCAATATTGGCAATAATGATGT GTCC
168	SGSTSNIGN HYVS	351	TCTGGAAGCACCTCCAACATTGGGAATCATTATGT ATCC
169	SGSSSNIGD NDVS	352	TCAGGAAGCAGCTCCAATATTGGGGATAATGATGT ATCC
170	SGYSSNIGN NYVS	353	TCTGGATACAGCTCCAACATTGGGAATAATTATGT ATCC
171	SGSGSNIGN NFVS	354	TCTGGAAGCGGCTCCAACATTGGAAATAATTTTGT ATCC
172	SGSSSNIW NNYVS	355	TCTGGAAGCAGCTCCAACATTTGGAATAATTATGT ATCC
173	FGSSSNIGN NYVS	356	TTTGGAAGCAGCTCCAACATTGGGAATAATTATGT ATCC
174	SGSSSNIEK NYVS	357	TCTGGAAGCAGCTCCAACATTGAGAAGAATTATGT ATCC
175	SGSRSNIGN YYVS	358	TCTGGAAGTAGATCCAATATTGGAAATTATTATGT ATCC

176	SGTKSNIG NNYVS	359	TCTGGAACCAAGTCAAACATTGGGAATAATTATGT ATCT
177	SGSTSNIGN YYVS	360	TCTGGAAGCACCTCCAACATTGGGAATTATTATGT ATCC
178	SGTSSNIGN NYVA	361	TCTGGAACCAGCTCCAACATTGGGAATAATTATGT GGCC
179	PGTSSNIGN NYVS	362	CCTGGAACCAGCTCCAACATTGGGAATAATTATGT ATCC
180	SGSTSNIGI NYVS	363	TCCGGAAGCACCTCCAACATTGGGATTAATTATGT ATCC
181	SGSSSNIGS NLVS	364	TCTGGAAGCAGCTCCAACATTGGGAGTAATCTGGT ATCC
182	SGSSSNIGN NHVS	365	TCTGGAAGCAGCTCCAACATTGAGAATAATCATGT ATCC
183	SGTRSNIGN NYVS	366	TCTGGAACCAGGTCCAACATCGGCAATAATTATGT TTCG
184	SGSTSNIGD NYVS	367	TCTGGAAGCACCTCCAACATTGGGGACAATTATGT TTCC
185	SGGSSNIGK NFVS	368	TCTGGAGGCAGTTCCAACATTGGGAAGAATTTTGT ATCC
186	SGSRSDIGN NYVS	369	TCTGGAAGCAGGTCCGACATTGGGAATAATTATGT ATCC
187	SGTSSNIGN NDVS	370	TCTGGAAGTAGCTCCAACATTGGGAATAATGATGT ATCC
188	SGSSSNIGS KYVS	371	TCTGGAAGCAGCTCCAACATTGGGAGTAAATATGT ATCA
189	SGSSFNIGN NYVS	372	TCTGGAAGCAGCTTCAACATTGGGAATAATTATGT ATCC
190	SGSSSNIGN TYVS	373	TCTGGAAGCAGCTCCAACATTGGGAATACTTATGT ATCC
191	SGSSSNIGD NHVS	374	TCTGGAAGCAGCTCCAATATTGGGGATAATCATGT ATCC
192	SGSSSNIGN NHVS	375	TCTGGAAGCAGCTCCAACATTGGCAATAATCATGT TTCC
193	SGSTSNIGN NDVS	376	TCTGGAAGCACCTCCAACATTGGGAATAATGATGT ATCC
194	SGSRSNVG NNYVS	377	TCTGGAAGCAGATCCAACGTTGGCAATAATTATGT TTCA
195	SGGTSNIG KNYVS	378	TCCGGAGGCACCTCCAACATTGGGAAGAATTATGT GTCT
196	SGSSSNIAD NYVS	379	TCTGGAAGCAGCTCCAACATTGCCGATAATTATGT TTCC
197	SGSSSNIGA NYVS	380	TCTGGAAGCAGCTCCAACATTGGCGCCAATTATGT ATCC
198	SGSSSNIGS NYVA	381	TCTGGAAGCAGCTCCAACATTGGGAGTAATTATGT GGCC
199	SGSSSNIGN NFLS	382	TCTGGAAGCAGCTCCAACATTGGGAACAATTTTCT CTCC
200	SGRSSNIGK NYVS	383	TCTGGAAGAAGCTCCAACATTGGGAAGAATTATGT ATCC

201	SGSSPNIGA NYVS	384	TCTGGAAGCAGCCCCAACATTGGGGCTAATTATGT ATCC
202	SGSSSNIGP NYVS	385	TCCGGAAGCAGCTCCAACATTGGGCCTAATTATGT GTCC
203	SGSSSTIGN NYIS	386	TCTGGAAGCAGCTCCACCATTGGGAATAATTATAT ATCC
204	SGSSSNIGN YFVS	387	TCTGGAAGCAGCTCCAACATTGGGAATTATTTTGT ATCC
205	SGSRSNIGN NFVS	388	TCTGGAAGCCGCTCCAACATTGGTAATAATTTTGT TCC
206	SGGSSNIGS NFVS	389	TCTGGAGGCAGCTCCAACATTGGGAGTAATTTTGT ATCC
207	SGSSSNIGY NYVS	390	TCTGGAAGCAGCTCCAACATTGGGTATAATTATGT ATCC
208	SGTSSNIEN NYVS	391	TCTGGAACCAGCTCGAACATTGAGAACAATTATGT ATCC
209	SGSSSNIGN YYVS	392	TCTGGAAGTAGCTCCAACATTGGGAATTATTATGT ATCC
210	SGSTSNIGK NYVS	393	TCTGGAAGCACCTCCAACATTGGGAAGAATTATGT ATCC
211	SGSSSNIGT YYVS	394	TCTGGAAGCAGTTCCAACATTGGGACTTATTATGTC TCT
212	SGSSSNVG KNYVS	395	TCTGGAAGCAGCTCCAACGTTGGGAAAAATTATGT ATCT
213	SGSTSNIGD NFVS	396	TCTGGAAGCACCTCCAACATTGGGGATAATTTTGT ATCC
214	SGSTSNIGT NYVS	397	TCTGGAAGCACCTCCAACATTGGAACTAATTATGT TTCC
215	SGGTSNIG NNYVS	398	TCTGGAGGTACTTCCAACATTGGGAATAATTATGT CTCC
216	SGSYSNIGN NYVS	399	TCTGGAAGCTACTCCAATATTGGGAATAATTATGT ATCC
217	SGSSSNIED NYVS	400	TCTGGAAGCAGCTCCAACATTGAAGATAATTATGT ATCC
218	SGSSSNIGK HYVS	401	TCTGGAAGCAGCTCCAACATTGGGAAACATTATGT ATCC
219	SGSGSNIGS NYVS	402	TCCGGTTCCGGCTCAAACATTGGAAGTAATTATGT CTCC
220	SGSSSNIGN NYIS	403	TCTGGAAGCAGCTCCAACATTGGAAATAATTATAT ATCA
221	SGASSNIGN NYVS	404	TCTGGAGCCAGTTCCAACATTGGGAATAATTATGT TTCC
222	SGRTSNIGN NYVS	405	TCTGGACGCACCTCCAACATCGGGAACAATTATGT ATCC
223	SGGSSNIGS NYVS	406	TCTGGAGGCAGCTCCAATATTGGGAGTAATTACGT ATCC
224	SGSGSNIGN NYVS	407	TCTGGAAGCGGCTCCAACATTGGGAATAATTATGT ATCC
225	SGSTSNIGS NYVS	408	TCTGGAAGCACCTCCAACATTGGGAGTAATTATGT ATCC

226	SGSSSIGN NYVA	409	TCTGGAAGCAGCTCCAGCATTGGGAATAATTATGT GGCG
227	SGSSNLG NNYVS	410	TCTGGAAGCAGTTCCAACCTTGGAAATAATTATGT ATCC
228	SGTSSNIGK NYVS	411	TCTGGAACCAGCTCCAACATTGGGAAAAAATTATGT ATCC
229	SGSSSDIGN KYIS	412	TCTGGAAGCAGCTCCGATATTGGGAACAAGTATAT ATCC
230	SGSSSNIGS NYIS	413	TCTGGAAGCAGCTCCAACATTGGAAGTAATTACAT ATCC
231	SGSTSNIGA NYVS	414	TCTGGAAGCACCTCCAACATTGGGGCTAACTATGT GTCC
232	SGSSSNIGN KYVS	415	TCTGGAAGCAGCTCCAACATTGGGAATAAGTATGT ATCC
233	SGSSSNIGN NYGS	416	TCTGGAAGCAGCTCCAACATTGGGAATAATTATGG ATCC
234	SGSTSNIAN NYVS	417	TCTGGAAGCACCTCCAACATTGCGAATAATTATGT ATCC
235	SGSYSNIGS NYVS	418	TCTGGAAGCTACTCCAATATTGGGAGTAATTATGT ATCC
236	SGSSSNIGS NFVS	419	TCTGGAAGCAGCTCCAACATTGGGAGTAATTTTGT ATCC
237	SGSSSNLEN NYVS	420	TCTGGAAGCAGCTCCAATCTTGAGAATAATTATGT ATCC
238	SGSISNIGS NYVS	421	TCTGGAAGCATCTCCAATATTGGCAGTAATTATGT ATCC
239	SGSSSDIGS NYVS	422	TCTGGAAGCAGCTCCGACATTGGGAGTAATTATGT ATCC
240	SGSSSNIGT NYVS	423	TCTGGAAGCAGCTCCAACATTGGGACTAATTATGT ATCC
241	SGSSSNIGK NFVS	424	TCTGGAAGCAGCTCCAACATTGGGAAGAATTTTGT ATCC
242	SGSSSNIGN NFIS	425	TCTGGAAGCAGCTCCAACATTGGGAATAATTTTAT ATCC
243	SGGSSNIGN NYVS	426	TCTGGAGGCAGCTCCAACATTGGCAATAATTATGT TTCC
244	SGSSSNIGE NYVS	427	TCTGGAAGCAGCTCCAACATTGGGGAGAATTATGT ATCC
245	SGSSSNIGN NFVA	428	TCTGGAAGCAGCTCCAATATTGGGAATAATTTTGT GGCC
246	SGGSSNIGN NYVA	429	TCTGGAGGCAGCTCCAACATTGGGAATAATTATGT AGCC
247	SGSSSHIGN NYVS	430	TCTGGAAGCAGCTCCCACATTGGAAATAATTATGT ATCC
248	SGSSSNIGS NDVS	431	TCTGGAAGCAGCTCCAATATTGGAAGTAATGATGT ATCG
249	SGSSSNIGN NYVT	432	TCTGGAAGCAGCTCCAACATTGGGAATAATTATGT AACC
250	SGSSSNIGN NPVS	433	TCTGGAAGCAGCTCCAACATTGGGAATAATCCTGT ATCC

251	SGSSNIGN HYVS	434	TCTGGAGGCAGCTCCAATATTGGGAATCATTATGT ATCC
252	SGTSSNIGN NYVS	435	TCTGGAACCAGCTCCAACATTGGGAATAATTATGT ATCC
253	SGSSSNIGS NYVS	436	TCTGGAAGCAGCTCCAACATTGGAAGTAATTATGT CTCG
254	SGGTSNIGS NYVS	437	TCTGGAGGCACCTCCAACATTGGAAGTAATTATGT ATCC
255	SGSKSNIGN NYVS	438	TCTGGAAGCAAGTCCAACATTGGGAATAATTATGT ATCC
256	SGRSSNIGN NYVS	439	TCTGGAAGAAGCTCCAACATTGGGAATAATTATGT ATCG
257	SGSSSNVG SNYVS	440	TCTGGAAGCAGCTCCAACGTTGGGAGTAATTATGT TTCC
258	SGSTSNIGN NFVS	441	TCTGGAAGCACCTCCAATATTGGGAATAATTTTGT ATCC
259	SGSNFNIGN NYVS	442	TCTGGAAGCAACTTCAACATTGGGAATAATTATGT CTCC
260	SGSTSNIGY NYVS	443	TCTGGAAGCACCTCCAATATTGGATATAATTATGT ATCC
261	SGSSSNIVS NYVS	444	TCTGGAAGCAGCTCCAATATTGTAAGTAATTATGT ATCC
262	SGTSSNIGN NFVS	445	TCTGGAACCAGCTCCAACATTGGGAATAATTTTGT ATCC
263	SGSSSNIGR NFVS	446	TCTGGAAGCAGCTCCAACATTGGGAGGAATTTTGT GTCC
264	SGTTSNIGN NYVS	447	TCTGGAACGACCTCCAACATTGGGAATAATTATGT CTCC
265	SGSSSNIGN NDVS	448	TCTGGAAGCAGCTCCAACATTGGGAATAATGATGT ATCC
266	SGSSSNIGN HDVS	449	TCTGGAAGCAGCTCCAACATTGGGAATCATGATGT ATCC
267	SGSSSNIGS SHVS	450	TCTGGAAGCAGCTCCAACATTGGAAGTAGTCATGT ATCC
268	SGSSSNIGI HYVS	451	TCTGGAAGCAGCTCCAACATTGGGATTCATTATGT ATCC
269	SGGGSNIG YNYVS	452	TCTGGAGGCGGCTCCAACATTGGCTATAATTATGT CTCC
270	SGSSSNIGD HYVS	453	TCTGGAAGCAGCTCCAACATTGGGGATCATTATGT GTCG
271	SGSSSNLG KNYVS	454	TCTGGAAGCAGCTCCAACCTTGGGAAGAATTATGT ATCT
272	SGSSSNIGD NFVS	455	TCTGGAAGCAGCTCCAACATTGGCGATAATTTTGT ATCC
273	SGSTSNIEK NYVS	456	TCTGGAAGCACCTCCAACATTGAGAAAACTATGT ATCG
274	SGSSSNIGK DYVS	457	TCTGGAAGCAGCTCCAACATTGGGAAGGATTATGT ATCC
275	SGSSSNIGK NYVS	458	TCTGGAAGCAGCTCCAACATTGGGAAGAATTATGT ATCC

276	SGSSSNIGN NYVS	459	TCTGGAAGCAGCTCCAACATTGGGAATAATTATGT ATCC
277	SGSSSNIGN NYAS	460	TCTGGAAGCAGCTCCAACATTGGGAATAATTATGC CTCC
278	SGISSNIGN NYVS	461	TCTGGAATCAGCTCCAACATTGGGAATAATTATGT ATCC
279	TGSSSNIGN NYVS	462	ACTGGAAGCAGCTCCAACATTGGGAATAATTATGT ATCC
280	SGTSSNIGN NHVS	463	TCTGGAACCAGCTCCAACATTGGGAATAATCATGT TTCC
281	SGSRSNIGK NYVS	464	TCTGGAAGTCGTTCCAACATTGGGAAAAATTATGT ATCC
IGLV1-51-L2			
465	DNNKRPP	616	GACAATAATAAGCGACCCCCA
466	ENNRPS	617	GAGAATAATAGGCGACCCTCA
467	DNNKQPS	618	GACAATAATAAGCAACCCTCA
468	DNNKRPL	619	GACAATAACAAGCGACCCTTG
469	DNDKRPA	620	GACAATGATAAGCGACCCGCA
470	DNHERPS	621	GACAATCATGAGCGACCCTCA
471	ENRKRPS	622	GAAAACCGTAAGCGACCCTCA
472	DNDQRPS	623	GACAATGATCAGCGACCCTCA
473	ENYKRPS	624	GAGAATTATAAGCGACCCTCA
474	ENTKRPS	625	GAAAATACTAAGCGACCCTCA
475	DTEKRPS	626	GACACTGAGAAGAGGCCCTCA
476	DNDKRPP	627	GACAATGATAAGCGACCCCCA
477	DHNRKPS	628	GACCATAATAAGCGACCCTCA
478	GNNRPS	629	GGCAATAATGAGCGACCCTCA
479	DTSKRPS	630	GACACTAGTAAGCGACCCTCA
480	EYNKRPS	631	GAATATAATAAGCGCCCCTCA
481	ENIKRPS	632	GAAAATATTAAGCGACCCTCA
482	DNVKRPS	633	GACAATGTTAAGCGACCCTCA
483	ENDKRSS	634	GAAAACGATAAACGATCCTCA
484	ENNRHS	635	GAAAATAATAAGCGACACTCA
485	GNDQRPS	636	GGAAATGATCAGCGACCCTCA
486	DNDRPS	637	GACAATGATAGGCGACCCTCA
487	DNHKRPS	638	GACAATCATAAGCGGCCCTCA
488	DNNDRPS	639	GACAATAATGACCGACCCTCA
489	ENNQRPS	640	GAGAATAATCAGCGACCCTCA
490	DNNQRPS	641	GACAATAATCAGCGACCCTCA
491	ENVKRPS	642	GAGAATGTTAAGCGACCCTCA
492	DTYKRPS	643	GACACTTATAAGAGACCCTCA
493	NNNNRPS	644	AACAATAATAACCGACCCTCA
494	GNNNRPS	645	GGCAATAATAATCGACCCTCA
495	ENDQRPS	646	GAAAATGATCAGCGACCCTCA
496	DNNKRAS	647	GACAATAATAAGCGAGCCTCA
497	DNDKRPL	648	GACAATGATAAGCGACCCTTA
498	DTDERPS	649	GACACTGATGAGCGACCTTCA

499	DNRKRPS	650	GACAATAGGAAGCGACCCTCA
500	DNDARPS	651	GACAATGATGCTCGACCCTCA
501	DNNKRLS	652	GACAATAATAAGCGACTCTCA
502	DNDKRAS	653	GACAATGATAAGCGAGCCTCA
503	DNTERPS	654	GACAATACTGAGCGACCCTCA
504	DNNIRPS	655	GACAATAATATTCGACCCTCA
505	DNKRPS	656	GACAATAAGAGGCGACCCTCA
506	DDNNRPS	657	GACGATAATAACCGACCCTCA
507	ANNRRPS	658	GCGAATAATCGACGACCCTCA
508	DNDKRLS	659	GACAATGATAAGCGACTGTCA
509	DNNKRPA	660	GACAATAATAAGCGACCCGCA
510	DNYRRPS	661	GACAATTATAGACGTCCCTCA
511	ANDQRPS	662	GCCAATGATCAGCGACCCTCA
512	DNDKRRS	663	GACAATGATAAGCGACGCTCA
513	DKNERPS	664	GACAAGAATGAGCGACCCTCA
514	DNKERPS	665	GACAATAAGGAGCGACCCTCA
515	DNNKGPS	666	GACAATAATAAGGGACCCTCA
516	ENDRRPS	667	GAAAATGATAGACGACCCTCA
517	ENDERPS	668	GAAAATGATGAGCGACCCTCA
518	QNNKRPS	669	CAAAATAATAAGCGACCCTCA
519	DNRERPS	670	GACAATCGTGAGCGACCCTCA
520	DNNRRPS	671	GACAATAATAGACGACCCTCA
521	GNNRRPS	672	GGAAATAATAGGCGACCCTCA
522	DNDNRPS	673	GACAATGATAACCGACCCTCA
523	EDNKRPS	674	GAAGATAATAAGCGACCCTCA
524	DDDERPS	675	GACGATGATGAGCGGCCCTCA
525	ASNKRPS	676	GCAAGTAATAAGCGACCCTCA
526	DNNKRSS	677	GACAATAATAAGCGATCCTCA
527	QNNERPS	678	CAAAATAATGAGCGACCCTCA
528	DDDRRPS	679	GACGATGATAGGCGACCCTCA
529	NNDKRPS	680	AACAATGATAAGCGACCCTCA
530	DNNNRPS	681	GACAATAATAACCGACCCTCA
531	DNNVRPS	682	GACAATAATGTGCGACCCTCA
532	ENNERPS	683	GAAAATAATGAGCGACCCTCA
533	DNNHRPS	684	GACAATAATCACCGACCCTCA
534	DNDERPS	685	GACAATGATGAGCGCCCCTCG
535	DNIRRPS	686	GACAATATCCGGCGACCCTCA
536	DFNKRPS	687	GACTTTAATAAGCGACCCTCA
537	ETNKRPS	688	GAAACTAATAAGCGACCCTCA
538	NDNKRPS	689	AACGATAATAAGCGACCCTCA
539	DDNKRPS	690	GACGATAATAAGCGACCCTCA
540	DNYKRPS	691	GACAATTATAAGCGACCCTCA
541	HNNKRPS	692	CACAATAATAAGCGACCCTCA
542	DNHQRP	693	GACAATCATCAGCGACCCTCA
543	DNYKRAS	694	GACAATTATAAGCGAGCCTCA
544	DNIKRPS	695	GACAATATTAAGCGACCCTCA

545	DTHKRPS	696	GACACTCATAAGCGACCCTCA
546	DTNRRPS	697	GACACTAATAGGCGACCCTCT
547	DTNQRPS	698	GACACTAATCAGCGACCCTCA
548	ESDKRPS	699	GAAAGTGATAAGCGACCCTCA
549	DNDKRSS	700	GACAATGATAAGCGATCTTCG
550	GSNKRPS	701	GGCAGTAATAAGCGACCCTCA
551	DNNKRVS	702	GACAATAACAAGCGAGTTTCA
552	NNNRRPS	703	AACAATAATAGGCGACCCTCA
553	DNFKRPS	704	GACAATTTTAAAGCGACCCTCA
554	ENDKRPS	705	GAAAATGATAAACGACCCTCA
555	ENNKRLS	706	GAAAATAATAAGCGACTCTCA
556	ADNKRPS	707	GCAGATAATAAGCGACCCTCA
557	EDNERPS	708	GAAGATAATGAGCGCCCCTCA
558	DTDQRPS	709	GACACTGATCAGCGACCCTCA
559	DNYQRPS	710	GACAATTATCAGCGACCCTCA
560	DENKRPS	711	GACGAGAATAAGCGACCCTCA
561	DTNKRPS	712	GACACTAATAAGCGACCCTCA
562	DDYRRPS	713	GACGATTATCGGCGACCCTCA
563	DNDKRHS	714	GACAACGATAAGCGGCACTCA
564	ENDNRPS	715	GAAAATGATAATCGACCCTCA
565	DDNERPS	716	GACGATAATGAGCGCCCCTCA
566	DNKKRPS	717	GACAATAAGAAGCGACCCTCA
567	DVDKRPS	718	GACGTTGATAAGCGACCCTCA
568	ENKKRPS	719	GAAAATAAAAAACGACCCTCT
569	VNDKRPS	720	GTCAATGATAAGCGACCCTCA
570	DNDHRPS	721	GACAATGATCACCGACCCTCA
571	DINKRPS	722	GACATTAATAAGCGACCCTCA
572	ANNERPS	723	GCCAATAATGAGCGACCCTCA
573	DNENRPS	724	GACAATGAAAACCGACCGTCA
574	GDDKRPS	725	GGCGATGATAAGCGACCCTCA
575	ANNQRPS	726	GCCAATAATCAGCGACCTTCA
576	DDDKRPS	727	GACGATGATAAGCGACCCTCA
577	YNNKRPS	728	TACAATAATAAGCGGCCCTCA
578	EDDKRPS	729	GAAGATGATAAGCGACCCTCA
579	ENNNRPS	730	GAAAACAATAACCGACCCTCG
580	DNNLRPS	731	GACAATAATCTGCGACCCTCA
581	ESNKRPS	732	GAGAGTAACAAGCGACCCTCA
582	DTDKRPS	733	GACACTGATAAGCGGCCCTCA
583	DDDQRPS	734	GACGATGATCAGCGACCCTCA
584	VNNKRPS	735	GTGAATAATAAGAGACCCTCC
585	DDYKRPS	736	GACGATTATAAGCGACCCTCA
586	DNTKRPS	737	GACAATACTAAGCGACCCTCA
587	DDTERPS	738	GACGATACTGAGCGACCCTCA
588	GNDKRPS	739	GGCAATGATAAGCGACCCTCA
589	DNEKRPS	740	GACAATGAAAAGCGACCCTCA
590	DNDDRPS	741	GACAATGATGACCGACCCTCA

591	DDNRRPS	742	GACGATAATAGGCGTCCCTCA
592	GNNKRPS	743	GGCAATAATAAGCGACCCTCA
593	ANDKRPS	744	GCCAATGATAAGCGACCCTCA
594	DNNKRHS	745	GACAATAATAAGCGACACTCA
595	DDNQRP	746	GACGACAATCAGCGACCCTCA
596	GNDRRPS	747	GGCAATGATAGGCGACCCTCA
597	DNHNRRPS	748	GACAATCATAACCGACCCTCA
598	DNYERPS	749	GACAATTATGAGCGACCCTCA
599	ENNKRSS	750	GAAAAATAATAAGCGATCCTCA
600	DDHKRPS	751	GACGATCATAAGCGGCCCTCA
601	DNNKRPS	752	GACAATAATAAACGACGTTCA
602	DNDKRPS	753	GACAATGATAAGCGACCGTCA
603	DKNKRPS	754	GACAAGAATAAGCGACCCTCA
604	DNNKRPS	755	GACAATAATAAGCGACCCTCA
605	DIDKRPS	756	GACATTGATAAGCGACCCTCA
606	DDKKRPS	757	GACGATAAGAAGCGACCCTCA
607	ANNKRPS	758	GCCAATAATAAGCGACCCTCA
608	DNDKGPS	759	GACAATGATAAGGGACCCTCA
609	EDNRRPS	760	GAAGATAATAGGCGACCCTCA
610	ENNKRPS	761	GAGAATAATAAGCGACCCTCA
611	NNNKRPS	762	AACAATAATAAGCGACCCTCA
612	DNNRPS	763	GACAATAATGAGCGACCCTCA
613	DNIQRPS	764	GACAATATTGAGCGACCCTCA
614	DNNYRPS	765	GACAATAATTACCGACCCTCA
615	DNYNRPS	766	GACAATTATAACCGACCCTCA
IGLV1-51-L3			
767	CGTWDTSLSAVVF	1591	TGCGGAACATGGGATACCAGCCTGAGTGCTGTGGTGTTC
768	CGTWDTSLSAGVF	1592	TGCGGAACATGGGATACCAGCCTGAGTGCTGGGGTGTTC
769	CGTWDTSLSAWVF	1593	TGCGGAACATGGGATACCAGCCTGAGTGCTTGGGTGTTC
770	CGTWDRSLSAGVF	1594	TGCGGAACATGGGATAGGAGCCTGAGTGCGGGGGTGTTC
771	CGTWDRSLSAWVF	1595	TGCGGAACATGGGATAGGAGCCTGAGTGCTTGGGTATT
772	CGTWDTSLSGGVF	1596	TGCGGAACATGGGATACCAGCCTGAGTGGTGGGGTGTTC
773	CGTWDTSLSRAGVF	1597	TGCGGAACATGGGATACTAGCCTGCGTGCTGGCGTCTTC
774	CGTWDRSLSVWVF	1598	TGCGGAACATGGGATAGGAGCCTGAGTGTTTGGGTGTTC
775	CGTWDTSLSVVVF	1599	TGCGGAACATGGGATACCAGTCTGAGTGTTGTGGTCTTC
776	CGTWDTSLSAAVF	1600	TGCGGAACGTGGGATACCAGCCTGAGTGCTGCGGTGTTC
777	CGAWDTSLSAGVF	1601	TGCGGAGCATGGGATACCAGCCTGAGTGCTGGAGTGTTC

778	CATWDTSL SAVVF	1602	TGCGCAACATGGGATACCAGCCTGAGTGCTGTGGT ATTC
779	CATWDTSL SAGVF	1603	TGCGCAACATGGGATACCAGCCTGAGTGCTGGTGT GTTC
780	CGTWESSL SAWVF	1604	TGTGGAACATGGGAGAGCAGCCTGAGTGCTTGGGT GTTC
781	CGTWDCTL SAGVF	1605	TGCGGAACATGGGATACCACCCTGAGTGCGGGTGT CTTC
782	CGTWDCTL SVWVF	1606	TGCGGAACATGGGATACTAGCCTGAGTGTGTGGGT GTTC
783	CGTWDCTL SVGVF	1607	TGCGGAACATGGGATACTAGCCTGAGTGTGTGGGT GTTC
784	CGTWDCTL STGVF	1608	TGCGGAACATGGGACACCAGTCTGAGCACTGGCGT CTTC
785	CGTWDCTL SGVVF	1609	TGCGGAACATGGGATACCAGCCTGAGTGGTGTGGT CTTC
786	CGTWDCTL SAYVF	1610	TGCGGAACATGGGATACCAGCCTGAGTGCTTATGT CTTC
787	CGTWDCTL SAEVF	1611	TGCGGAACATGGGATACCAGCCTGAGTGCTGAGGT GTTC
788	CGTWDCTL SAGVF	1612	TGCGGAACATGGGATACCGCCTGAGTGCTGGGGT ATTC
789	CGTWDRSL SAYVF	1613	TGCGGAACATGGGATAGGAGCCTGAGTGCTTATGT CTTC
790	CGTWDRSL SAVVF	1614	TGCGGAACATGGGATAGGAGCCTCAGTGCCGTGGT ATTC
791	CGTWDCTL SAWVF	1615	TGCGGAACATGGGATAACACCCTGAGTGCGTGGGT GTTC
792	CGTWDNR LSAGVF	1616	TGCGGAACATGGGATAACAGGCTGAGTGCTGGGGT GTTC
793	CGTWDISL SAWVF	1617	TGCGGAACATGGGACATCAGCCTGAGTGCTTGGGT GTTC
794	CGTWHSSL SAGVF	1618	TGCGGAACATGGCATAGCAGCCTGAGTGCTGGGGT ATTC
795	CGTWGSSL SAWVF	1619	TGCGGAACATGGGGTAGCAGTTTGAGTGCTTGGGT GTTC
796	CGTWESSL SGWVF	1620	TGCGGAACATGGGAGAGCAGCCTGAGTGGTTGGGT GTTC
797	CGTWESSL SAVVF	1621	TGCGGAACATGGGAGAGCAGCCTGAGTGCTGTGGT TTTC
798	CGTWDYSL SAVVF	1622	TGCGGAACATGGGATTACAGCCTGAGTGCTGTGGT ATTC
799	CGTWDYSL SAGVF	1623	TGCGGAACATGGGATTACAGCCTGAGTGCTGGGGT ATTC
800	CGTWDVSL SVGVF	1624	TGCGGAACATGGGATGTCAGCCTGAGTGTGTGGAGT GTTC
801	CGTWDCTL SAVVF	1625	TGCGGAACATGGGATACCACCCTGAGTGCTGTGGT TTTC
802	CGTWDCTL NIGVF	1626	TGCGGAACATGGGATACCACTCTGAATATTGGGGT GTTC

803	CGTWDTSL TAVVF	1627	TGCGGAACATGGGATACCAGCCTGACTGCTGTGGT ATTC
804	CGTWDTSL TAAVF	1628	TGCGGAACCTGGGATACCAGCCTGACTGCTGCTGT GTTC
805	CGTWDTSL SVGLF	1629	TGCGGCACATGGGATACCAGCCTGAGTGTGGGGCT ATTC
806	CGTWDTSL SGRVF	1630	TGCGGAACCTGGGATACCAGCCTGAGTGGTAGGGT GTTC
807	CGTWDTSL SGAVF	1631	TGCGGAACATGGGATACCAGCCTGAGTGGTGCAGT GTTC
808	CGTWDTSL SAGLF	1632	TGCGGAACATGGGATACCAGCCTGAGTGCTGGCCT GTTC
809	CGTWDTSL SAGGVF	1633	TGCGGAACATGGGATACCAGCCTGAGTGCTGGAGG GGTCTTC
810	CGTWDTSL RAYVF	1634	TGCGGAACATGGGATACCAGCCTGCGTGCTTATGT CTTC
811	CGTWDTSL RAWVF	1635	TGCGGAACATGGGATACTAGTTTGCGTGCTTGGGT ATTC
812	CGTWDTSL NTGVF	1636	TGCGGAACATGGGATACCAGCCTGAATACTGGGGT ATTC
813	CGTWDTSL NIWVF	1637	TGCGGAACATGGGATACCAGCCTGAATATTTGGGT GTTC
814	CGTWDTSL NIGVF	1638	TGCGGAACATGGGATACAAGCCTGAATATTGGGGT GTTC
815	CGTWDTSL IAVVF	1639	TGCGGAACATGGGATACCAGCCTGATTGCTGTGGT GTTC
816	CGTWDRLS SGWVF	1640	TGCGGAACGTGGGATAGGAGCCTGAGTGGTTGGGT GTTC
817	CGTWDNR LSGWVF	1641	TGCGGAACATGGGATAACAGGCTGAGTGGTTGGGT GTTC
818	CGTWDKSL SAVVF	1642	TGCGGAACGTGGGATAAGAGCCTGAGTGCTGTGGT CTTC
819	CGTWDKG LSAWVF	1643	TGCGGAACATGGGATAAAGGCCTGAGTGCTTGGGT GTTC
820	CGTWDISL SAGVF	1644	TGCGGAACATGGGATATCAGCCTGAGTGCTGGGGT GTTC
821	CGTWDLSL SGGEVVF	1645	TGCGGAACATGGGATGAGAGCCTGAGTGGTGGCG AGGTGGTCTTC
822	CGTWDASL SAWVF	1646	TGCGGAACATGGGATGCCAGCCTGAGTGCTGGGT GTTC
823	CGTWDAG LSAWVF	1647	TGCGGAACATGGGATGCCGGCCTGAGTGCTTGGGT GTTC
824	CGAWDTSL SAWVF	1648	TGCGGAGCATGGGATACCAGCCTGAGTGCTTGGGT GTTC
825	CGAWDTSL SAVVF	1649	TGCGGAGCATGGGATACCAGCCTGAGTGCTGTGGT GTTC
826	CGAWDTSL RAGVF	1650	TGCGGAGCATGGGATACCAGCCTGCGTGCTGGGGT TTTC
827	CATWDTSV SAWVF	1651	TGCGCAACATGGGATACCAGCGTGAGTGCTTGGGT GTTC

828	CATWDTSL SAWVF	1652	TGCGCAACATGGGATACCAGCCTGAGTGCGTGGGT GTTC
829	CATWDNTL SAGVF	1653	TGCGCAACATGGGACAACACCCTGAGTGCTGGGGT GTTC
830	CAAWDRSL SVWVF	1654	TGCGCAGCATGGGATAGGAGCCTGAGTGTTTGGGT GTTC
831	CYTWHSSL RGGVF	1655	TGCTACACATGGCATTCCAGTCTGCGTGTTGGGGT GTTC
832	CVTWTSSP SAWVF	1656	TGCGTAACGTGGACTAGTAGCCCCGAGTGCTTGGGT GTTC
833	CVTWRGG LVLF	1657	TGCGTGACATGGCGTGGTGGCCTTGTGTGTTC
834	CVTWDTSL TSVVL	1658	TGCGTAACATGGGATACCAGCCTGACTTCTGTGGT ACTC
835	CVTWDTSL SVYWVF	1659	TGCGTAACATGGGATACCAGCCTGAGTGTTTATTG GGTGTTC
836	CVTWDTSL SAWVF	1660	TGCGTTACATGGGATACCAGCCTGAGTGCCTGGGT GTTC
837	CVTWDTDL SVALF	1661	TGCGTCACATGGGATACCGACCTCAGCGTTGCGCT CTTC
838	CVTWDRSL SGWVF	1662	TGCGTAACATGGGATAGGAGCCTGAGTGGTTGGGT GTTC
839	CVTWDRSL REVLF	1663	TGCGTAACATGGGATCGCAGCCTGAGAGAGGTGTT ATTCTTC
840	CVTWDRSL RAVVF	1664	TGCGTAACATGGGATCGCAGCCTGAGAGCGGTGGT ATTCTTC
841	CVTWDRSL DAGVF	1665	TGCGTAACATGGGACAGGAGCCTCGATGCTGGGGT TTTC
842	CVTWDNTL SAGVF	1666	TGCGTGACATGGGATAACACCCTGAGTGCTGGGGT CTTC
843	CVTWDNN LFGVVF	1667	TGCGTAACATGGGATAACAACCTGTTTGGTGTGGT CTTC
844	CVSWDTSL SGAVF	1668	TGCGTATCATGGGATACCAGCCTGAGTGGTGCGGT ATTCTTC
845	CVSWDTSL SAGVF	1669	TGCGTCTCATGGGATACCAGCCTGAGTGCTGGGGT ATTCTTC
846	CTTWFRTP SDVVF	1670	TGCACAACATGGGTTTAGGACTCCGAGTGATGTGGT CTTC
847	CTTWFRTA SDVVF	1671	TGCACAACATGGGTTTAGGACTGCGAGTGATGTGGT CTTC
848	CTTWDYGL SVVF	1672	TGCACAACGTGGGATTACGGTCTGAGTGTCGTCTT C
849	CTARDTSL SPGGVF	1673	TGCACAGCAAGGGATACCAGCCTGAGTCCTGGCGG GGTCTTC
850	CSTWNTRP SDVVF	1674	TGCTCAACATGGAATACGAGGCCGAGTGATGTGGT GTTC
851	CSTWESSL TTVVF	1675	TGTTCAACATGGGAGAGCAGTTTGACTACTGTGGT CTTC
852	CSTWDTSL TNVLF	1676	TGCTCAACATGGGATACCAGCCTCACTAATGTGCT ATTCTTC

853	CSTWDTSL SGVVF	1677	TGCTCAACATGGGATACCAGCCTGAGTGGAGTAGT CTTC
854	CSTWDHSL KAALF	1678	TGCTCAACATGGGATCACAGCCTGAAAGCTGCACT GTTC
855	CSTWDARL SVRVF	1679	TGCTCAACCTGGGATGCGAGGCTGAGTGTCCGGGT GTTC
856	CSSYTSSST WVF	1680	TGCTCCTCATATACAAGCAGCAGCACTTGGGTGTT C
857	CSSYATRG LRVLF	1681	TGCAGCTCATACGCAACCCGCGGCCTTCGTGTGTT GTTC
858	CSSWDATL SVRIF	1682	TGTTTCATCATGGGACGCCACCCTGAGTGTTCGCAT ATTC
859	CQVWEGSS DHWVF	1683	TGTCAGGTGTGGGAGGGTAGTAGTGATCATTGGGT GTTC
860	CQTWDNR LSAVVF	1684	TGCCAAACCTGGGATAACAGACTGAGTGCTGTGGT GTTC
861	CQTWDHSL HVGVF	1685	TGTCAAACGTGGGATCACAGCCTGCATGTTGGGGT GTTC
862	CQSYDDIL NVWVL	1686	TGCCAGTCCATGACGACATCTTGAATGTTTGGGTC CTT
863	CNTWDKSL TSELF	1687	TGCAATACATGGGATAAGAGTTTGACTTCTGAACT CTTC
864	CLTWDRSL NVRVF	1688	TGCTTAACATGGGATCGCAGCCTGAATGTGAGGGT GTTC
865	CLTWDHSL TAYVF	1689	TGCCTAACATGGGACCACAGCCTGACTGCTTATGT CTTC
866	CLTRDTSLS APVF	1690	TGCTTAACAAGGGATACCAGTCTGAGTGCCCTGT GTTC
867	CKTWESGL NFGHVF	1691	TGCAAAACATGGGAAAGTGGCCTTAATTTTGGCCA CGTCTTC
868	CKTWDTS� SAVVF	1692	TGCAAAACATGGGATACCAGCCTGAGTGCTGTGGT CTTC
869	CGVWDVS LGAGVF	1693	TGCGGAGTCTGGGATGTCAGTCTGGGTGCTGGGGT GTTC
870	CGVWDTPP SAVLF	1694	TGCGGAGTCTGGGATACCACCCCGAGTGCCGTTCT TTTC
871	CGVWDTTL SAVLF	1695	TGCGGAGTCTGGGATACCACCCTGAGTGCCGTTCT TTTC
872	CGVWDTS� GVF	1696	TGCGGAGTATGGGATACCAGCCTGGGGGTCTTC
873	CGVWDTN LGKWVF	1697	TGCGGGGTATGGGATACCAACCTGGGTAAATGGGT TTTC
874	CGVWDTG LDAGWVF	1698	TGTGGAGTTTGGGATACTGGCCTGGATGCTGGTTG GGTGTTC
875	CGVWDNV LEAYVF	1699	TGCGGAGTGTGGGATAACGTCCTGGAGGCCTATGT CTTC
876	CGVWDISL SANWVF	1700	TGCGGAGTCTGGGATATCAGCCTGAGTGCTAATTG GGTGTTC
877	CGVWDHS LGIWAF	1701	TGCGGAGTATGGGATCACAGCCTGGGGATTTGGGC CTTC

878	CGVWDDIL TAEVF	1702	TGCGGAGTTTGGGATGATATTCTGACTGCTGAAGT GTTC
879	CGVRDTSL GVF	1703	TGCGGAGTTCGGGATACCAGCCTGGGGGTCTTC
880	CGTYDTSL PAWVF	1704	TGCGGAACATACGATACGAGCCTGCCTGCTTGGGT GTTT
881	CGTYDNLV FGYVF	1705	TGCGGAACTTACGATAATCTTGTATTTGGTTATGTC TTC
882	CGTYDDRL REVF	1706	TGCGGAACATACGATGATAGACTCAGAGAGGTGTT C
883	CGTWVTSL SAGVF	1707	TGCGGAACGTGGGTTACCAGCCTGAGTGCTGGGGT GTTC
884	CGTWVSSL TTVVF	1708	TGCGGAACATGGGTTAGCAGCCTGACTACTGTAGT ATTC
885	CGTWVSSL NVWVF	1709	TGCGGAACATGGGTTAGCAGCCTGAACGTCTGGGT GTTC
886	CGTWVGRF WVF	1710	TGCGGAACATGGGTTGGCAGGTTTTGGGTATTC
887	CGTWSGGP SGHWLF	1711	TGCGGAACATGGTCTGGCGGCCCGAGTGGCCATTG GTTGTTC
888	CGTWSGGL SGHWLF	1712	TGCGGAACATGGTCTGGCGGCCTGAGTGGCCATTG GTTGTTC
889	CGTWQTG REAVLF	1713	TGCGGAACGTGGCAGACCGGCCGGGAGGCTGTCCT ATTT
890	CGTWQSRL RWVF	1714	TGCGGAACGTGGCAGAGCAGGCTGAGGTGGGTGTT C
891	CGTWQSRL GWVF	1715	TGCGGAACGTGGCAGAGCAGGCTGGGGTGGGTGTT C
892	CGTWPRSL SAWVF	1716	TGCGGAACATGGCCTAGGAGCCTGAGTGCTGTTTG GGTGTTC
893	CGTWNLY LSAGDVVF	1717	TGCGGAACATGGAATAACTACCTGAGTGCTGGCGA TGTGGTTTTTC
894	CGTWLGSQ SPYWVF	1718	TGCGGAACATGGCTTGGCAGCCAGAGTCCTTATTG GGTCTTC
895	CGTWHTGL SAYVF	1719	TGCGGAACATGGCATAACCGGCCTGAGTGCTTATGT CTTC
896	CGTWHSTL SAGHWVF	1720	TGCGGAACATGGCATAGTACCCTGAGTGCTGGCCA TTGGGTGTTC
897	CGTWHSSL STWVF	1721	TGCGGAACATGGCATAGTAGCCTGAGTACTTGGGT GTTC
898	CGTWHSSL SAYVF	1722	TGCGGAACATGGCATAGCAGCCTGAGTGCTATGT CTTC
899	CGTWHSSL SAVVF	1723	TGCGGAACATGGCATAGCAGCCTGAGTGCTGTGGT ATTC
900	CGTWHSSL SGWVF	1724	TGCGGAACGTGGCATTCGGGCCTGAGTGGGTGGGT TTTC
901	CGTWHNTL RNVIF	1725	TGCGGAACATGGCATAACACCCTGCGTAATGTGAT ATTC
902	CGTWHASL TAVF	1726	TGCGGAACATGGCATGCCAGCCTGACTGCTGTGTT C

903	CGTWGWY GSQRGVVF	1727	TGCGGGACATGGGGATGGTATGGCAGCCAGAGAG GCGTCGTCTTC
904	CGTWGWY GGQRGVVF	1728	TGCGGGACATGGGGATGGTATGGCGGCCAGAGAG GCGTCGTCTTC
905	CGTWGTSL SAWVF	1729	TGCGGAACCTGGGGAACCAGCCTGAGTGCTTGGGT GTTC
906	CGTWGSSL TTGLF	1730	TGCGGAACCTGGGGTAGCAGCCTGACTACTGGCCT GTTC
907	CGTWGSSL TAYVF	1731	TGCGGAACATGGGGTAGCAGCCTGACTGCCTATGT CTTC
908	CGTWGSSL SVVF	1732	TGCGGAACATGGGGTAGCAGCCTGAGTGTTGTGTT C
909	CGTWGSSL SGGVF	1733	TGCGGAACATGGGGTAGCAGCCTGAGTGGTGGGGT GTTC
910	CGTWGSSL SAYWVF	1734	TGCGGAACATGGGGTAGCAGCCTGAGTGCTTATTG GGTGTTC
911	CGTWGSSL SAYVVF	1735	TGCGGAACATGGGGTAGCAGCCTGAGTGCTTATGT GGTGTTC
912	CGTWGSSL SAYVF	1736	TGCGGAACATGGGGTAGCAGCCTGAGTGCTTATGT CTTC
913	CGTWGSSL SAVVF	1737	TGCGGAACGTGGGGTAGTAGCCTGAGTGCTGTGGT GTTC
914	CGTWGSSL SAPYVF	1738	TGCGGAACATGGGGTAGCAGCCTGAGTGCTCCTTA TGTCTTC
915	CGTWGSSL SAPVF	1739	TGCGGAACATGGGGTAGCAGCCTGAGTGCCCCGGT GTTC
916	CGTWGSSL SAGVF	1740	TGCGGAACATGGGGTAGCAGCCTGAGTGCTGGGGT GTTC
917	CGTWGSSL SAGLF	1741	TGCGGAACTTGGGGTAGCAGCCTGAGTGCTGGACT GTTC
918	CGTWGSSL SAGALF	1742	TGCGGAACATGGGGTAGCAGCCTGAGTGCTGGGGC ACTCTTC
919	CGTWGSSL RAWVF	1743	TGCGGAACATGGGGCAGTAGCCTGCGTGCTTGGGT GTTC
920	CGTWFTSL ASGVF	1744	TGCGGAACCTGGTTTACTAGTCTGGCTAGTGGGGT TTTC
921	CGTWETSL SVVVI	1745	TGCGGAACTTGGGAGACCAGTCTGAGTGTCGTGGT CATC
922	CGTWETSL SGVF	1746	TGCGGAACATGGGAGACCAGCCTGAGTGGTGTCTT C
923	CGTWETSL SDWVF	1747	TGCGGAACATGGGAAACCAGCCTGAGTGATTGGGT ATTC
924	CGTWETSL SAGVF	1748	TGCGGAACATGGGAGACCAGCCTGAGTGCTGGGGT ATTC
925	CGTWETSL NYVAF	1749	TGCGGAACATGGGAAACCAGCCTTAATTATGTGGC CTTC
926	CGTWETSL NTWLL	1750	TGCGGAACATGGGAGACCAGCCTGAATACTTGTT GCTC
927	CGTWETSE SGNYIF	1751	TGCGGAACATGGGAGACCAGCGAGAGTGGTAATT ACATCTTC

928	CGTWETRL GTWVI	1752	TGCGGAACATGGGAAACCAGACTGGGTACTTGGGT GATC
929	CGTWETQL YWVF	1753	TGCGGAACATGGGAGACCCAGTTATATTGGGTGTT C
930	CGTWETGL SAGEVF	1754	TGCGGAACATGGGAGACTGGCCTAAGTGCTGGAGA GGTGTTC
931	CGTWESTL SVFLF	1755	TGCGGAACTTGGGAAAGCACCTGAGTGTTTTCTT ATTC
932	CGTWESSL TVVVF	1756	TGCGGGACATGGGAAAGTAGCCTGACTGTTGTGGT CTTC
933	CGTWESSL TGVVF	1757	TGCGGAACATGGGAAAGTAGCCTGACTGGAGTGGT ATTC
934	CGTWESSL TGFVF	1758	TGCGGAACATGGGAAAGCAGCCTGACTGGTTTTGT CTTC
935	CGTWESSL SVGVF	1759	TGTGGAACATGGGAGAGCAGCCTGAGTGTTGGGGT GTTC
936	CGTWESSL SEWVF	1760	TGCGGAACCTGGGAAAGTAGCCTCAGTGAATGGGT GTTC
937	CGTWESSL SAVF	1761	TGCGGAACATGGGAGAGCAGCCTGAGTGCTGTATT C
938	CGTWESSL SAGYIF	1762	TGCGGAACATGGGAGAGCAGCCTGAGTGCTGGTTA TATCTTC
939	CGTWESSL SAGVF	1763	TGCGGAACATGGGAGAGCAGCCTGAGTGCTGGAGT GTTC
940	CGTWESSL SAGPVF	1764	TGCGGAACATGGGAAAGCAGCCTGAGCGCTGGCCC GGTGTTC
941	CGTWESSL SAGGQVF	1765	TGCGGAACATGGGAAAGCAGCCTGAGTGCTGGAG GCCAGGTGTTC
942	CGTWESSL SAFGGYVF	1766	TGCGGAACATGGGAGAGCAGCCTGAGTGCCCTTCGG CGGTTATGTCTTC
943	CGTWESSL RVWVF	1767	TGCGGAACATGGGAAAGCAGCCTGAGGGTTTGGGT GTTC
944	CGTWESSL FTGPWVF	1768	TGCGGAACATGGGAAAGCAGCCTCTTTACTGGGCC TTGGGTGTTC
945	CGTWESLS ATYVF	1769	TGCGGAACATGGGAGAGCCTGAGTGCCACCTATGT CTTC
946	CGTWESGL SAGVF	1770	TGCGGAACATGGGAGAGCGGCCTGAGTGCTGGTGT CTTC
947	CGTWESDF WVF	1771	TGCGGAACATGGGAAAGCGACTTTTGGGTGTTT
948	CGTWENRL SAVVF	1772	TGCGGTACATGGGAAAACAGACTGAGTGCTGTGGT CTTC
949	CGTWENRL SAGVF	1773	TGCGGAACATGGGAAAACAGACTGAGTGCCGGGG TATTC
950	CGTWEISL TTSVVF	1774	TGCGGAACATGGGAAATCAGCCTGACTACTTCTGT GGTATTC
951	CGTWEISLS TSVVF	1775	TGCGGAACATGGGAAATCAGCCTGAGTACTTCTGT GGTATTC
952	CGTWESGL SVVF	1776	TGCGGAACATGGGAAGGCAGCCTCAGTGTTGTTTT C

953	CGTWECSL RVF	1777	TGCGGAACATGGGAAGGCAGCCTGAGGGTGTTC
954	CGTWECSL RHVF	1778	TGCGGAACATGGGAGGGCAGCCTGAGGCACGTGTT C
955	CGTWDYSP VRAGVF	1779	TGCGGAACATGGGATTACAGCCCTGTACGTGCTGG GGTGTTC
956	CGTWDYSL SVYLF	1780	TGCGGAACGTGGGATTACAGCCTGAGTGTTCCTC
957	CGTWDYSL SSGVVF	1781	TGCGGAACATGGGATTACAGCCTGAGTTCCTGGCGT GGTATTC
958	CGTWDYSL SAWVF	1782	TGCGGAACATGGGATTACAGCCTGAGTGCCTGGGT GTTC
959	CGTWDYSL SAEVF	1783	TGCGGAACATGGGATTACAGTCTGAGTGCTGAGGT GTTC
960	CGTWDYSL RRAIF	1784	TGCGGAACATGGGATTACAGCCTGCGTCGTGCGAT ATTC
961	CGTWDWS LILQLF	1785	TGCGGAACATGGGATTGGAGCCTCATTCTCAATT GTTC
962	CGTWDVTL HTGVF	1786	TGCGGAACATGGGATGTCACCTTGACACTGGGGT GTTC
963	CGTWDVTL HIGVF	1787	TGCGGAACATGGGATGTCACCTTGACATATTGGGGT GTTC
964	CGTWDVTL HAGVF	1788	TGCGGAACATGGGATGTCACCTTGACATGCTGGGGT GTTC
965	CGTWDVSL YSGGVF	1789	TGCGGAACATGGGATGTCAGTTTGTATAGTGGCGG GGTCTTC
966	CGTWDVSL TSFVF	1790	TGTGGAACATGGGATGTCAGCCTGACTTCTTCGTC TTC
967	CGTWDVSL SVGVL	1791	TGCGGAACATGGGATGTCAGCCTGAGTGTGGGGT GCTC
968	CGTWDVSL SAGDVVF	1792	TGCGGAACGTGGGATGTCAGCCTGAGTGCTGGCGA TGTAAGTTTC
969	CGTWDVSL NVVVF	1793	TGCGGAACATGGGATGTCAGCCTGAATGTCGTGGT TTTC
970	CGTWDVSL NTQVF	1794	TGCGGAACATGGGATGTCAGCCTGAATACTCAGGT GTTC
971	CGTWDVSL GALF	1795	TGCGGCACATGGGATGTGAGCCTGGGTGCGCTGTT C
972	CGTWDVN LKTUVF	1796	TGCGGAACGTGGGACGTTAATCTGAAAACGTGCTG TTTC
973	CGTWDVIL SAEVF	1797	TGCGGAACATGGGATGTCATCCTGAGTGCTGAGGT ATTC
974	CGTWDTTV SAVVF	1798	TGCGGAACATGGGATACACCGTGAGTGCTGTGGT TTTC
975	CGTWDTTL TAWVF	1799	TGCGGAACATGGGATACACCCCTGACTGCCTGGGT GTTC
976	CGTWDTTL SVFLF	1800	TGCGGAACATGGGACACCACCTTGAGTGTTTTCTC ATTC
977	CGTWDTSV SAGVF	1801	TGCGGGACTTGGGATACAGTGTGAGTGCTGGGGT GTTC

978	CGTWDTSV ISWVF	1802	TGCGGAACATGGGATACCAGTGTGATTTCTTGGGT TTTC
979	CGTWDTSR SSLYVVF	1803	TGCGGAACATGGGATACCAGTCGGAGTTCTCTCTA TGTGGTCTTC
980	CGTWDTSR SAWVF	1804	TGCGGAACATGGGATACCAGCCGGAGTGCTTGGGT ATTC
981	CGTWDTSR NPGGIF	1805	TGCGGAACATGGGATACCAGCCGGAATCCTGGAGG AATTTTC
982	CGTWDTSR GHVF	1806	TGCGGAACATGGGACACCAGTCGGGGTCATGTTTT C
983	CGTWDTSR STGQVLF	1807	TGCGGAACATGGGATACCAGCCCGAGTACTGGCCA GGTGCTTTTC
984	CGTWDTSR SAWVF	1808	TGCGGAACATGGGATACCAGCCCGAGTGCTTGGGT GTTC
985	CGTWDTSR TWVF	1809	TGCGGAACATGGGATACTAGCCTGACCTGGGTGTT C
986	CGTWDTSR TWFAVF	1810	TGCGGAACATGGGATACCAGCCTGACGTGGTTCGC AGTGTTTC
987	CGTWDTSR TVVVF	1811	TGCGGAACATGGGATACCAGCCTGACTGTTGTGGT ATTC
988	CGTWDTSR TTSWVF	1812	TGCGGAACATGGGATACCAGCCTGACTACTTCTTG GGTGTTTC
989	CGTWDTSR TTGPFVVF	1813	TGCGGAACATGGGATACCAGCCTGACCACTGGTCC TTTTTGGGTGTTTC
990	CGTWDTSR TPFYVF	1814	TGCGGAACATGGGATACCAGCCTGACTCCTTTTTAT GTCTTC
991	CGTWDTSR TAYVF	1815	TGCGGAACATGGGATACCAGCCTGACTGCTTATGT CTTC
992	CGTWDTSR TAWVF	1816	TGCGGAACATGGGATACCAGCCTGACTGCTTGGGT GTTC
993	CGTWDTSR TAWGVF	1817	TGCGGAACATGGGATACCAGCCTGACTGCGTGGGG GGTGTTTC
994	CGTWDTSR TAVVL	1818	TGCGGCACATGGGATACCAGCCTGACTGCGGTGGT TCTC
995	CGTWDTSR TARVF	1819	TGCGGAACCTGGGATACCAGCCTGACTGCTCGGGT TTTC
996	CGTWDTSR TAIVF	1820	TGCGGAACATGGGATACCAGCCTGACTGCGATTGT CTTC
997	CGTWDTSR TAGVF	1821	TGCGGAACATGGGATACCAGCCTGACTGCTGGGT CTTC
998	CGTWDTSR SVYVF	1822	TGCGGAACATGGGATACCAGCCTGAGTGTTTATGT CTTC
999	CGTWDTSR SVVF	1823	TGCGGAACATGGGATACCAGCCTGAGTGTTGGTGT C
1000	CGTWDTSR SVGEF	1824	TGCGGGACATGGGATACCAGCCTGAGTGTTGGGGA ATTC
1001	CGTWDTSR STWVF	1825	TGCGGAACATGGGATACCAGCCTGAGTACTTGGGT GTTC
1002	CGTWDTSR STVVF	1826	TGCGGAACATGGGATACCAGCCTGAGTACTGTGGT ATTC

1003	CGTWDTSLS STGQVLF	1827	TGCGGAACATGGGATACCAGCCTGAGTACTGGCCA GGTGCTTTTC
1004	CGTWDTSLS STGPLWVF	1828	TGCGGCACATGGGATACCAGCCTGAGCACTGGTCC TCTTTGGGTGTTC
1005	CGTWDTSLS SSYVF	1829	TGCGGAACTTGGGATACCAGCCTGAGTTCTTATGT CTTC
1006	CGTWDTSLS SSVVF	1830	TGCGGAACATGGGATACCAGCCTGAGTTCTGTGGT CTTC
1007	CGTWDTSLS SSRYIF	1831	TGCGGAACATGGGATACCAGCCTGAGTTCTAGATA CATATTC
1008	CGTWDTSLS SSRFIF	1832	TGCGGAACATGGGATACCAGCCTGAGTTCTAGATT CATATTC
1009	CGTWDTSLS SSGWVF	1833	TGCGGAACATGGGATACCAGCCTGAGTTCTGGGTG GGTGTTTC
1010	CGTWDTSLS SRYVF	1834	TGCGGAACATGGGATACCAGCCTGAGTCGGTATGT GTTC
1011	CGTWDTSLS SQWLF	1835	TGCGGAACTTGGGATACCAGTCTGAGTCAATGGCT GTTC
1012	CGTWDTSLS SPGLWVF	1836	TGCGGAACATGGGATACCAGCCTGAGTCCTGGCCT TTGGGTGTTC
1013	CGTWDTSLS SNYVF	1837	TGCGGAACATGGGATACCAGCCTGAGTAATTATGT CTTC
1014	CGTWDTSLS SIWVF	1838	TGCGGAACATGGGATACCAGCCTAAGTATTTGGGT GTTC
1015	CGTWDTSLS SIGPFWVF	1839	TGCGGCACATGGGATACCAGCCTGAGCATTGGTCC TTTTTGGGTGTTC
1016	CGTWDTSLS SGWVF	1840	TGCGGAACATGGGATACCAGCCTGAGTGTTGGGT GTTC
1017	CGTWDTSLS SGTVF	1841	TGCGGAACATGGGATACCAGCCTGAGTGGTACAGT GTTC
1018	CGTWDTSLS SGGQVF	1842	TGCGGAACATGGGATACTAGTCTGAGTGGTGGCCA GGTGTTTC
1019	CGTWDTSLS SGGIF	1843	TGCGGAACATGGGATACCAGCCTGAGTGGTGGGAT ATTC
1020	CGTWDTSLS SGEDVVI	1844	TGCGGAACATGGGATACCAGCCTGAGTGGTGAGGA TGTGGTAATC
1021	CGTWDTSLS SFLYAF	1845	TGCGGAACATGGGATACCAGCCTGAGTTTCCTTTA TGCTTTTC
1022	CGTWDTSLS SEVVF	1846	TGCGGAACATGGGATACCAGCCTGAGTGAGGTCGT ATTC
1023	CGTWDTSLS SEVF	1847	TGCGGAACATGGGATACCAGCCTGAGTGAAGTGTT C
1024	CGTWDTSLS SENWVF	1848	TGCGGAACATGGGATACTAGCCTGAGTGAAAATTG GGTGTTTC
1025	CGTWDTSLS SAYIF	1849	TGCGGAACATGGGATACCAGCCTGAGTGCCTACAT ATTC
1026	CGTWDTSLS SAVVL	1850	TGCGGAACATGGGATACCAGCCTGAGTGCTGTGGT ACTC
1027	CGTWDTSLS SAVF	1851	TGCGGAACATGGGATACCAGCCTGAGTGCTGTTTT C

1028	CGTWDTSLSARVF	1852	TGCGGAACATGGGATACCAGCCTGAGTGCCCGGGTGTTC
1029	CGTWDTSLSARQVF	1853	TGCGGCACATGGGATACCAGCCTGAGTGCCCGCCA GGTATTC
1030	CGTWDTSLSALVF	1854	TGCGGAACATGGGATACCAGCCTGAGTGCTTTGGT TTTC
1031	CGTWDTSLSAKVF	1855	TGCGGAACATGGGATACCAGCCTGAGTGCTAAGGT GTTC
1032	CGTWDTSLSAKIF	1856	TGCGGAACATGGGATACCAGCCTGAGTGCGAAAAT CTTC
1033	CGTWDTSLSAKAVF	1857	TGCGGAACATGGGATACCAGCCTGAGTGCCAAGGC GGTATTC
1034	CGTWDTSLSAHAVF	1858	TGCGGAACATGGGATACCAGCCTGAGTGCCCATGC TGTGTTC
1035	CGTWDTSLSAGYVF	1859	TGCGGAACATGGGATACCAGCCTGAGTGCTGGCTA TGTCTTC
1036	CGTWDTSLSAGRWF	1860	TGCGGAACATGGGACACCAGTCTGAGTGCTGGCCG CTGGGTGTTC
1037	CGTWDTSLSAGIF	1861	TGCGGAACATGGGATACCAGCCTGAGTGCTGGGAT ATTC
1038	CGTWDTSLSAGGFRVF	1862	TGCGGAACATGGGATACCAGCCTGAGTGCTGGTGG GTTCCGGGTCTTC
1039	CGTWDTSLSAGAF	1863	TGCGGAACATGGGATACCAGCCTGAGTGCTGGGGC ATTC
1040	CGTWDTSLSADWFF	1864	TGCGGAACATGGGATACCAGTCTGAGTGCTGATTG GTTTTTTC
1041	CGTWDTSLSADEYVF	1865	TGCGGAACATGGGATACCAGCCTGAGTGCTGATGA ATATGTCTTC
1042	CGTWDTSLSAAWVF	1866	TGCGGCACATGGGATACCAGCCTGAGTGCGGCTTG GGTGTTC
1043	CGTWDTSLSAALF	1867	TGCGGAACATGGGATACCAGCCTGAGTGCTGCGCT ATTC
1044	CGTWDTSLSAAGVF	1868	TGCGGAACATGGGATACCAGCCTGAGTGCTGCGGG GTTTTTC
1045	CGTWDTSLSRVVVF	1869	TGCGGAACATGGGATACCAGCCTGAGAGTTGTGGT TTTC
1046	CGTWDTSLSRTWVF	1870	TGCGGAACATGGGATACCAGCCTGAGAACCTGGGT ATTC
1047	CGTWDTSLSRGAVF	1871	TGCGGAACGTGGGATACCAGCCTGAGGGGTGCAGT GTTC
1048	CGTWDTSLSRAVVF	1872	TGCGGAACATGGGATACCAGCCTGCGTGCTGTGGT ATTC
1049	CGTWDTSLSNVVYVF	1873	TGCGGAACATGGGATACAAGCCTGAATGTAGTTTA TGTCTTC
1050	CGTWDTSLSNTYLF	1874	TGCGGAACATGGGATACCAGCCTCAACACCTACCT GTTC
1051	CGTWDTSLSNFAWLF	1875	TGCGGAACATGGGATACTAGCCTGAACCTTCGCTTG GCTGTTC
1052	CGTWDTSLSLVWLF	1876	TGCGGCACATGGGATACCAGCCTTCTTGTGTGGCTT TTC

1053	CGTWDTSL KTWVF	1877	TGCGGAACATGGGATACCAGTCTGAAGACGTGGGT GTTC
1054	CGTWDTSL IVWVF	1878	TGCGGAACATGGGATACCAGTCTGATTGTCTGGGT GTTC
1055	CGTWDTSL ITGVF	1879	TGCGGAACATGGGATACCAGCCTAATTACTGGGGT GTTC
1056	CGTWDTSL ISVVF	1880	TGCGGAACATGGGATACCAGCCTGATTAGCGTGGT ATTC
1057	CGTWDTSL IAYVF	1881	TGCGGAACATGGGATACCAGCCTGATTGCTTATGT CTTC
1058	CGTWDTSL HTELF	1882	TGCGGAACATGGGATACCAGCCTGCACACTGAGTT GTTC
1059	CGTWDTSL GSYVF	1883	TGCGGAACTTGGGATACCAGCCTGGGTTCTTATGT CTTC
1060	CGTWDTSL GSLWVF	1884	TGCGGAACATGGGATACCAGCCTGGGTTCTCTTTG GGTGTTC
1061	CGTWDTSL GSGVF	1885	TGCGGTACATGGGATACCAGCCTGGGTTCTGGGGT ATTC
1062	CGTWDTSL GGRGVF	1886	TGCGGAACTTGGGATACCAGTCTGGGTGGTAGAGG GGTCTTC
1063	CGTWDTSL GAWVF	1887	TGCGGAACATGGGATACCAGCCTGGGTGCTTGGGT GTTC
1064	CGTWDTSL GAVVF	1888	TGCGGAACATGGGATACCAGCCTGGGTGCCGTGGT ATTC
1065	CGTWDTSL GAGVF	1889	TGCGGAACATGGGATACCAGCCTGGGTGCTGGGGT ATTC
1066	CGTWDTSL GAGLF	1890	TGCGGAACATGGGATACCAGCCTGGGTGCTGGCCT ATTC
1067	CGTWDTSL DAVVF	1891	TGCGGAACATGGGATACCAGTCTGGATGCTGTGGT TTTC
1068	CGTWDTSL DAVLF	1892	TGCGGGACTTGGGATACCAGCCTGGATGCTGTGCT GTTC
1069	CGTWDTSL AWVF	1893	TGCGGAACATGGGATACCAGCCTGGCTTGGGTGTT C
1070	CGTWDTSL ATGLF	1894	TGCGGAACATGGGATACCAGCCTGGCGACTGGACT GTTC
1071	CGTWDTSL APVVF	1895	TGCGGGACATGGGATACCAGCCTGGCCCCTGTAGT CTTC
1072	CGTWDTRL TIVIF	1896	TGCGGAACATGGGACACCCGCCTGACTATTGTGAT CTTC
1073	CGTWDTRL SVWLF	1897	TGTGGAACATGGGACACCAGGCTGAGTGTTTGGCT GTTC
1074	CGTWDTRL SVG VF	1898	TGCGGAACGTGGGACACCAGACTGAGTGTTGGGGT TTTC
1075	CGTWDTRL STVIF	1899	TGCGGCACATGGGATACCAGACTGAGTACTGTAAT TTTC
1076	CGTWDTRL SSVVF	1900	TGCGGAACATGGGATACCCGCCTGAGTTCTGTGGT CTTC
1077	CGTWDTRL SIVVF	1901	TGCGGAACATGGGATACCCGCCTGAGTATTGTGGT TTTC

1078	CGTWDTRL SAYVVF	1902	TGCGGAACATGGGATACCAGACTGAGTGCCTATGT GGTATTC
1079	CGTWDTRL SAWVF	1903	TGCGGAACCTGGGACACCCGCCTGAGTGCCTGGGT GTTC
1080	CGTWDTRL SAVVF	1904	TGCGGAACATGGGATACCAGACTGAGTGCTGTGGT GTTC
1081	CGTWDTRL SAGLF	1905	TGCGGAACATGGGATACCCGCCTGAGTGCTGGGT GTTC
1082	CGTWDTRL SAGGVF	1906	TGCGGAACATGGGATACCAGACTGAGTGCTGGTGG GGTGTTC
1083	CGTWDTRL NVWLF	1907	TGCGGAACATGGGATACCAGATTGAATGTGTGGCT ATTC
1084	CGTWDTN REVVLL	1908	TGCGGAACATGGGATACCAACCGGGAAGTTGTGCT CCTC
1085	CGTWDTNL RAHVF	1909	TGCGGAACATGGGATACCAACCTGCGTGCCCATGT CTTC
1086	CGTWDTNL PAVVF	1910	TGCGGAACATGGGATACTAATCTGCCCCGCTGTAGT GTTC
1087	CGTWDTNL GGVF	1911	TGCGGAACATGGGACACCAATTTGGGTGGGGTGT C
1088	CGTWDTIV SIGVF	1912	TGCGGAACATGGGATACCATCGTGAGTATTGGGGT GTTC
1089	CGTWDTIL SAVVF	1913	TGCGGAACATGGGATACCATCCTGAGTGCGGTGGT GTTC
1090	CGTWDTIL SAEVF	1914	TGCGGCACATGGGATACCATCCTGAGTGCTGAGGT GTTC
1091	CGTWDTHL GVVF	1915	TGCGGAACATGGGATACCCACCTGGGTGTGGTTTT C
1092	CGTWDTGP SPHWLF	1916	TGCGGAACATGGGATACCGGCCCCGAGCCCTCATTG GCTGTTC
1093	CGTWDTGL TFGGVF	1917	TGCGGAACATGGGATACCGGCCTGACTTTTGGAGG CGTGTTC
1094	CGTWDTGL TAFVF	1918	TGCGGAACATGGGATACCGGCCTGACTGCTTTTGT CTTC
1095	CGTWDTGL SVWVF	1919	TGCGGAACATGGGATACCGGCCTGAGTGTTTGGGT GTTC
1096	CGTWDTGL STGIF	1920	TGCGGAACATGGGATACCGGCCTGAGTACTGGGAT TTTC
1097	CGTWDTGL SSLLF	1921	TGCGGAACATGGGATACCGGCCTGAGTTCCCTGCT CTTC
1098	CGTWDTGL SIVVF	1922	TGCGGAACGTGGGACACCGGCCTGAGTATTGTGGT GTTC
1099	CGTWDTGL SFVVF	1923	TGCGGAACGTGGGACACCGGCCTGAGTTTTGTGGT GTTC
1100	CGTWDTGL SAWVF	1924	TGCGGAACATGGGATACCGGCCTGAGTGCTTGGGT GTTC
1101	CGTWDTGL SAGVVF	1925	TGCGGAACATGGGATACCGGCCTGAGTGCTGGTGT GGTATTC
1102	CGTWDTGL RGWIF	1926	TGCGGAACATGGGATACCGGTCTGAGGGGTGGAT TTTC

1103	CGTWDTEL SAGVF	1927	TGCGGAACATGGGATACCGAGCTAAGTGCGGGGGT CTTC
1104	CGTWD TAL TAGVF	1928	TGCGGAACGTGGGATACCGCCCTGACTGCTGGGGT GTTC
1105	CGTWD TAL SLVVF	1929	TGCGGAACATGGGATACTGCCCTGAGTCTTGTGGT CTTC
1106	CGTWD TAL SAWLF	1930	TGCGGAACATGGGATACCGCCCTGAGTGCCTGGCT GTTC
1107	CGTWD TAL SAGVF	1931	TGCGGCACATGGGATACCGCCCTGAGTGTCTGGGGT GTTC
1108	CGTWD TAL RGVLF	1932	TGCGGAACATGGGATACCGCCCTGCGTGGCGTGCT GTTC
1109	CGTWD TAL KEWLF	1933	TGCGGAACATGGGATACCGCCCTGAAAGAATGGCT GTTC
1110	CGTWDRTL TAGDVLF	1934	TGCGGAACATGGGATAGGACCCTGACTGCTGGCGA TGTGCTCTTC
1111	CGTWDRSV TYVF	1935	TGCGGAACATGGGATAGAAGCGTGACTTATGTCTT C
1112	CGTWDRSR NEWVF	1936	TGCGGAACATGGGATCGCAGCCGAAATGAATGGGT GTTC
1113	CGTWDRSL TVWVF	1937	TGCGGAACATGGGATCGCAGTCTGACTGTTTGGGT CTTC
1114	CGTWDRSL TPGWLF	1938	TGCGGAACATGGGATCGCAGCCTGACTCCTGGGTG GTTGTTC
1115	CGTWDRSL TAWVF	1939	TGCGGAACATGGGATAGAAGCCTGACTGCTTGGGT GTTC
1116	CGTWDRSL SVVVF	1940	TGCGGAACATGGGACCGCAGCCTGAGTGTTGTGGT ATTC
1117	CGTWDRSL SVVF	1941	TGCGGCACATGGGATCGCAGCCTGAGTGTAAGTCTT C
1118	CGTWDRSL SVQLF	1942	TGCGGAACATGGGATAGGAGCCTGAGTGTTCAATT GTTC
1119	CGTWDRSL SVLWVF	1943	TGCGGAACATGGGATCGCAGCCTCAGTGTTCTTTG GGTGTTC
1120	CGTWDRSL SVGLF	1944	TGCGGAACATGGGATCGCAGCCTGAGTGTTGGATT ATTC
1121	CGTWDRSL STWVF	1945	TGCGGAACATGGGATCGCAGCCTGAGTACTTGGGT GTTC
1122	CGTWDRSL STHWVL	1946	TGCGGAACATGGGATAGAAGCCTGAGTACTCATTG GGTGCTC
1123	CGTWDRSL STHWVF	1947	TGCGGAACATGGGATAGAAGCCTGAGTACTCATTG GGTGTTC
1124	CGTWDRSL SSAVF	1948	TGCGGAACCTGGGATCGAAGCCTGAGTTCTGCGGT GTTC
1125	CGTWDRSL SPSYVF	1949	TGCGGAACATGGGACAGAAGCCTGAGTCCCTCTTA TGTCTTC
1126	CGTWDRSL SGEVF	1950	TGCGGAACATGGGATAGGAGCCTGAGTGGTGAGGT GTTC
1127	CGTWDRSL SGAVF	1951	TGCGGAACATGGGATAGGAGCCTGAGTGGTGCGGT GTTC

1128	CGTWDRSL SAVAF	1952	TGCGGAACATGGGATCGCAGCCTGAGTGCTGTGGC ATTC
1129	CGTWDRSL SAGGEF	1953	TGCGGAACATGGGATAGGAGCCTGAGTGCCGGGG GGGAATTC
1130	CGTWDRSL SAFWVF	1954	TGCGGAACATGGGATCGCAGCCTGAGTGCTTTTTG GGTGTTC
1131	CGTWDRSL SAAVF	1955	TGCGGAACATGGGATAGGAGCCTGAGTGCTGCGGT GTTC
1132	CGTWDRSL SAALF	1956	TGCGGAACATGGGATAGGAGCCTGAGTGCTGCACT CTTC
1133	CGTWDRSL RVF	1957	TGCGGAACATGGGATCGCAGCCTGAGAGTGTTC
1134	CGTWDRSL NWVF	1958	TGCGGTACATGGGACAGAAGCCTTAATTGGGTGTT C
1135	CGTWDRSL NVYVF	1959	TGCGGAACATGGGATCGCAGCCTGAATGTTTATGT CTTC
1136	CGTWDRSL NVGVF	1960	TGCGGAACATGGGATAGGAGCCTGAATGTTGGGGT GTTC
1137	CGTWDRSL HVVF	1961	TGCGGAACATGGGATCGGAGCCTGCATGTGGTCTT C
1138	CGTWDRSL GGWVF	1962	TGTGGAACATGGGATCGCAGCCTGGGTGGTTGGGT GTTC
1139	CGTWDRSL GAFWVF	1963	TGCGGAACATGGGATCGCAGCCTGGGTGCTTTTTG GGTGTTC
1140	CGTWDRSL FWVF	1964	TGCGGAACATGGGATAGAAGCCTGTTTTGGGTGTT C
1141	CGTWDRSL AAGVF	1965	TGCGGAACGTGGGATCGCAGCCTGGCTGCTGGGGT GTTC
1142	CGTWDRRL SGVVF	1966	TGCGGAACATGGGATAGGAGGTTGAGTGGTGTTCGT ATTC
1143	CGTWDRRL SDVVF	1967	TGCGGAACGTGGGATCGCCGCCTAAGTGATGTGGT ATTC
1144	CGTWDRRL SAVVF	1968	TGCGGAACATGGGATAGGAGGCTGAGTGCTGTGGT ATTC
1145	CGTWDRRL NVAFF	1969	TGCGGAACATGGGATAGACGCCTGAATGTTGCGTT CTTC
1146	CGTWDRRL LAVF	1970	TGTGGAACATGGGATAGGAGGCTGCTTGCTGTTTT C
1147	CGTWDRN LRAVVF	1971	TGCGGAACTTGGGATAGGAACCTGCGCGCCGTGGT CTTC
1148	CGTWDRLS AGVF	1972	TGCGGAACATGGGATAGGCTGAGTGCTGGGGTGTTC
1149	CGTWDRGP NTGVF	1973	TGCGGAACATGGGATAGAGGCCGAATACTGGGGT ATTC
1150	CGTWDRG LNTVYVF	1974	TGCGGAACATGGGATAGAGGCCTGAATACTGTTTA CGTCTTC
1151	CGTWDNY VSAPWVF	1975	TGCGGAACATGGGATAACTATGTGAGTGCCCCTTG GGTGTTC
1152	CGTWDNY LSAGDVVF	1976	TGCGGAACATGGGATAACTACCTGAGTGCTGGCGA TGTGGTTTTTC

1153	CGTWDNY LRAGVF	1977	TGCGGAACATGGGATAACTACCTGAGAGCTGGGGT CTTC
1154	CGTWDNY LGAVVF	1978	TGCGGAACATGGGACAATTATCTGGGTGCCGTGGT TTTC
1155	CGTWDNY LGAGVF	1979	TGCGGAACATGGGATAACTACCTGGGTGCGGGGGT GTTC
1156	CGTWDNT VSAPWVF	1980	TGCGGAACATGGGATAACACCGTGAGTGCCCCTTG GGTTTTTC
1157	CGTWDNTL SLWVF	1981	TGCGGAACATGGGATAACACCCTGAGTCTTTGGGT GTTC
1158	CGTWDNTL SAGVF	1982	TGCGGAACATGGGATAACACCCTGAGTGCTGGGGT CTTC
1159	CGTWDNTL LTVLF	1983	TGCGGAACATGGGACAACACTCTGCTTACTGTGTT ATTTC
1160	CGTWDNR LSSVIF	1984	TGCGGAACATGGGATAACAGACTGAGTAGTGTGAT TTTC
1161	CGTWDNR LSAVVF	1985	TGCGGAACATGGGATAACAGGTTGAGTGCTGTGGT CTTC
1162	CGTWDNR LSAGGIF	1986	TGCGGAACATGGGATAACAGGCTGAGTGCTGGTGG GATATTC
1163	CGTWDNR LSAEVF	1987	TGCGGAACATGGGATAACAGACTGAGTGCTGAGGT GTTC
1164	CGTWDNR LRGVLF	1988	TGTGGAACATGGGATAACAGACTGCGTGTTGGGGT TCTC
1165	CGTWDNR LLENVF	1989	TGCGGAACATGGGATAATCGCCTGCTTGAGAATGT CTTC
1166	CGTWDNN LRAVF	1990	TGCGGAACATGGGATAACAACCTGCGTGCTGTCTT C
1167	CGTWDNN LRAGVF	1991	TGCGGAACTTGGGATAATAACCTGCGTGCTGGAGT GTTC
1168	CGTWDNN LGGGRVF	1992	TGCGGAACATGGGACAACAATTTGGGCGGTGGCCG GGTGTTC
1169	CGTWDNN LGAGVL	1993	TGCGGAACATGGGATAACAACCTGGGTGCTGGCGT CCTC
1170	CGTWDNN LGAGVF	1994	TGCGGAACATGGGATAACAACCTGGGTGCTGGCGT CTTC
1171	CGTWDNIL SAAVF	1995	TGCGGAACTTGGGATAACATCCTGAGCGCTGCGGT GTTC
1172	CGTWDNIL DAGVF	1996	TGCGGAACCTGGGATAACATCTTGATGCAGGGGT TTTC
1173	CGTWDND LSGWLF	1997	TGCGGAACATGGGATAACGACCTGAGTGTTGGCT GTTC
1174	CGTWDND LSAWVF	1998	TGCGGAACATGGGATAACGACCTGAGTGCTGGGT GTTC
1175	CGTWDLTL GGVVF	1999	TGCGGAACATGGGATCTCACCTGGGTGGTGTGGT GTTC
1176	CGTWDLSL SAGVF	2000	TGCGGAACATGGGATCTCAGCCTGAGTGCTGGGGT ATTTC
1177	CGTWDLSL KEWVF	2001	TGCGGAACATGGGATCTCAGCCTGAAAGAATGGGT GTTC

1178	CGTWDLSL DAVVF	2002	TGCGGAACGTGGGATCTCAGCCTGGATGCTGTTGT TTTC
1179	CGTWDLK VF	2003	TGCGGAACCTGGGACCTGAAGGTTTTTC
1180	CGTWDKTL SVWVF	2004	TGCGGAACATGGGATAAGACTCTGAGTGTTTGGGT GTTC
1181	CGTWDKSL SVWVF	2005	TGCGGAACATGGGATAAGAGCCTGAGTGTTTGGGT GTTC
1182	CGTWDKSL SGVVF	2006	TGCGGAACATGGGATAAGAGCCTGAGTGGTGTGGT ATTT
1183	CGTWDKSL SDWVF	2007	TGCGGAACATGGGATAAGAGCCTGAGTGATTGGGT GTTC
1184	CGTWDKSL SALVF	2008	TGCGGAACATGGGATAAGAGCCTGAGTGCTTTGGT TTTC
1185	CGTWDKSL SAGVF	2009	TGCGGAACATGGGATAAGAGCCTGAGTGCTGGCGT CTTC
1186	CGTWDKSL SADV	2010	TGCGGAACATGGGATAAGAGCCTGAGTGCCGACGT CTTC
1187	CGTWDKR LTIVVF	2011	TGCGGAACATGGGATAAACGCCTGACTATTGTGGT CTTC
1188	CGTWDKR LSAWVL	2012	TGCGGAACATGGGATAAACGCCTGAGTGCTGGGT GCTC
1189	CGTWDKN LRAVVF	2013	TGCGGAACATGGGATAAGAACCTGCGTGCTGTGGT CTTC
1190	CGTWDITL SGFVF	2014	TGCGGAACATGGGATATCACCTGAGTGGGTTTGT CTTC
1191	CGTWDITL HTGVF	2015	TGCGGAACATGGGATATCACCTTGCATACTGGAGT ATTC
1192	CGTWDISV TVVF	2016	TGCGGAACATGGGATATCAGTGTGACTGTGGTGTT C
1193	CGTWDISV RGYAF	2017	TGCGGAACATGGGATATCAGTGTGAGGGGTTATGC CTTC
1194	CGTWDISR WVF	2018	TGCGGAACATGGGATATCAGCCGTTGGGTTTTTC
1195	CGTWDISP SAWVF	2019	TGCGGAACATGGGATATCAGCCCGAGTGCTTGGGT GTTC
1196	CGTWDISL SVWVF	2020	TGCGGAACATGGGATATTAGCCTGAGTGTCTGGGT GTTC
1197	CGTWDISL SVVF	2021	TGCGGAACATGGGATATCAGCCTGAGTGTTGTATT C
1198	CGTWDISL SSVVF	2022	TGCGGAACTTGGGATATCAGCCTGAGTTCTGTGGT GTTC
1199	CGTWDISL SHWLF	2023	TGCGGAACATGGGATATCAGCCTGAGTCACTGGTT GTTC
1200	CGTWDISL SGWVF	2024	TGCGGAACATGGGATATCAGTCTGAGTGGTTGGGT GTTC
1201	CGTWDISL SGRVF	2025	TGCGGAACATGGGATATCAGCCTGAGTGGTCGAGT GTTC
1202	CGTWDISL SAWAF	2026	TGCGGAACATGGGACATCAGCCTGAGTGCTTGGGC GTTC

1203	CGTWDISL SAVVF	2027	TGCGGAACATGGGATATCAGCCTGAGTGCTGTGGT TTTC
1204	CGTWDISL SAVIF	2028	TGCGGGACATGGGACATCAGCCTGAGTGCTGTGAT ATTTC
1205	CGTWDISL SAVF	2029	TGCGGAACATGGGATATCAGCCTGAGTGCTGTGTT C
1206	CGTWDISL SARVF	2030	TGCGGAACATGGGATATCAGCCTGAGTGCCCGGGT GTTC
1207	CGTWDISL SALVF	2031	TGCGGAACATGGGATATCAGCCTGAGTGCCCTGGT GTTC
1208	CGTWDISL SAHVF	2032	TGCGGAACATGGGATATTAGCCTGAGTGCCCATGT CTTC
1209	CGTWDISL SAGVVF	2033	TGCGGAACATGGGATATCAGCCTGAGTGCTGGGGT GGTATTTC
1210	CGTWDISL SAGPYVF	2034	TGCGGAACATGGGATATCAGCCTGAGTGCCGGCCC TTATGTCTTC
1211	CGTWDISL SAGGVF	2035	TGCGGCACATGGGATATCAGCCTGAGTGCTGGAGG GGTGTTC
1212	CGTWDISL SAEVF	2036	TGCGGAACATGGGATATCAGCCTGAGTGCTGAGGT TTTC
1213	CGTWDISL SAAVF	2037	TGCGGAACATGGGATATCAGCCTGAGTGCTGTCTGT GTTC
1214	CGTWDISL RAVF	2038	TGCGGAACATGGGATATCAGCCTGCGTGCTGTGTT C
1215	CGTWDISL NTGVF	2039	TGCGGAACATGGGATATTAGCCTGAATACTGGGGT GTTC
1216	CGTWDISL NNYVF	2040	TGCGGAACATGGGATATCAGCCTAAATAATTATGT CTTC
1217	CGTWDISLI AGVF	2041	TGCGGAACATGGGATATCAGCCTAATTGCTGGGGT ATTTC
1218	CGTWDISL HTWLF	2042	TGCGGAACATGGGATATCAGCCTGCATACTTGGCT GTTC
1219	CGTWDIRL TDELLF	2043	TGCGGAACATGGGATATCCGCCTGACCGATGAGCT GTTATTTC
1220	CGTWDIRL SGFVF	2044	TGCGGAACATGGGATATCAGACTGAGCGGTTTTGT TTTC
1221	CGTWDINL GAGGLYVF	2045	TGCGGAACATGGGATATCAACCTGGGTGCTGGGGG CCTTTATGTCTTC
1222	CGTWDIILS AEVF	2046	TGCGGAACATGGGATATCATCCTGAGTGCTGAGGT ATTTC
1223	CGTWDHTL SAVF	2047	TGCGGAACATGGGATCACACCCTGAGTGCTGTCTT C
1224	CGTWDHTL LTVLF	2048	TGCGGAACATGGGACCACACTCTGCTTACTGTGTT ATTTC
1225	CGTWDHSL TAVVF	2049	TGCGGAACATGGGATCACAGCCTGACTGCTGTGGT ATTTC
1226	CGTWDHSL TAGIF	2050	TGCGGAACCTGGGATCACAGCCTGACTGCTGGGAT ATTTC
1227	CGTWDHSL SVVLF	2051	TGCGGAACATGGGATCACAGCCTGAGTGTTGTATT ATTTC

1228	CGTWDHSL SLVF	2052	TGCGGAACATGGGATCACAGCCTGAGTTTGGTATT C
1229	CGTWDHSL SIGVF	2053	TGCGGAACATGGGATCACAGCCTGTCTATTGGGGT TTTC
1230	CGTWDHSL SAGVF	2054	TGCGGAACATGGGATCACAGCCTGAGTGCTGGGGT GTTC
1231	CGTWDHSL SAFVF	2055	TGTGGAACCTGGGATCACAGCCTGAGTGCTTTCGT GTTC
1232	CGTWDHSL SAAVF	2056	TGCGGAACATGGGATCACAGTCTGAGTGCTGCTGT TTTC
1233	CGTWDHN LRAVF	2057	TGCGGAACATGGGACCACAATCTGCGTGCTGTCTT C
1234	CGTWDFTL SVGRF	2058	TGCGGGACATGGGATTTCCACCCTGAGTGTTGGGCG CTTC
1235	CGTWDFTL SAPVF	2059	TGCGGAACATGGGATTTCCACCCTGAGTGCTCCTGT CTTC
1236	CGTWDFSV SAGWVF	2060	TGCGGAACGTGGGATTTCCAGCGTGAGTGCTGGGTG GGTGTTC
1237	CGTWDFSL TTWLF	2061	TGCGGAACGTGGGATTTCCAGTCTTACTACCTGGTTA TTC
1238	CGTWDFSL SVWVF	2062	TGCGGAACATGGGATTTCCAGCCTGAGTGTTTGGGT GTTC
1239	CGTWDFSL STGVF	2063	TGCGGAACATGGGATTTCCAGCCTGAGTACTGGGGT TTTC
1240	CGTWDFSL SGVVF	2064	TGCGGCACATGGGATTTCCAGCCTGAGTGGTGTGGT ATTC
1241	CGTWDFSL SGFVF	2065	TGCGGAACATGGGATTTCCAGCCTGAGTGGTTTCGT GTTC
1242	CGTWDFSL SAGVF	2066	TGCGGAACATGGGATTTCCAGCCTGAGTGCTGGGGT GTTC
1243	CGTWDETV RGWVF	2067	TGCGGAACATGGGATGAAACCGTGAGAGGTTGGGT GTTC
1244	CGTWDESL RSWVF	2068	TGCGGAACATGGGATGAAAGTCTGAGAAGCTGGGT GTTC
1245	CGTWDER QTDESYVF	2069	TGCGGAACCTGGGATGAGAGGCAGACTGATGAGTC CTATGTCTTC
1246	CGTWDERL VAGQVF	2070	TGCGGAACATGGGATGAGAGACTCGTTGCTGGCCA GGTCTTC
1247	CGTWDERL SPGAFF	2071	TGCGGAACATGGGATGAGAGACTGAGTCCTGGAGC TTTTTTC
1248	CGTWDEK VF	2072	TGCGGAACATGGGATGAGAAGGTGTTC
1249	CGTWDEG QTIDFFVF	2073	TGCGGAACCTGGGATGAAGGCCAGACTACTGATTT CTTTGTCTTC
1250	CGTWDDL AGVVF	2074	TGCGGAACATGGGATGACACCCTGGCTGGTGTGGT CTTC
1251	CGTWDDR LTSVF	2075	TGCGGAACATGGGATGACAGGCTGACTTCTGCGGT CTTC
1252	CGTWDDR LFVVVF	2076	TGCGGAACATGGGATGACAGACTGTTTGTGTGGT ATTC

1253	CGTWDDN LRGWVF	2077	TGCGGAACATGGGATGATAACCTGAGAGGTTGGGT GTTC
1254	CGTWDDN LRGVVF	2078	TGCGGAACATGGGATGACAACCTGCGTGGTGTCTG GTTC
1255	CGTWDDN LNIGRVF	2079	TGCGGAACCTGGGATGACAATTTGAATATTGGAAG GGTGTTC
1256	CGTWDDIL SAVIF	2080	TGCGGAACATGGGATGACATCCTGAGTGCTGTGAT ATTCTTC
1257	CGTWDDIL RGWVF	2081	TGCGGAACATGGGATGATATCCTGAGAGGTTGGGT GTTC
1258	CGTWDATL SPGWLF	2082	TGCGGAACATGGGATGCCACCCTGAGTCCTGGGTG GTTATTCTTC
1259	CGTWDAS VTSWVF	2083	TGCGGAACATGGGATGCCAGCGTGACTTCTTGGGT GTTC
1260	CGTWDASL TSVVF	2084	TGCGGAACATGGGATGCCAGCCTGACTTCTGTGGT CTTC
1261	CGTWDASL SVWVF	2085	TGCGGAACATGGGATGCCAGCCTGAGTGTTTGGGT GTTC
1262	CGTWDASL SVPWVF	2086	TGCGGAACATGGGATGCCAGCCTGAGTGTTCTTG GGTGTTC
1263	CGTWDASL SVAVF	2087	TGCGGAACATGGGATGCCAGCCTGAGTGTTGGCGGT ATTCTTC
1264	CGTWDASL STWVF	2088	TGCGGAACATGGGATGCCAGCCTGAGTACCTGGGT ATTCTTC
1265	CGTWDASL SGVVF	2089	TGCGGAACATGGGATGCCAGCCTGAGTGGTGTGGT ATTCTTC
1266	CGTWDASL SGGGEF	2090	TGCGGAACATGGGATGCCAGCCTGAGTGGTGGGGG AGAATTCTTC
1267	CGTWDASL SAGVF	2091	TGCGGAACATGGGATGCCAGCCTGAGTGCTGGGGT GTTC
1268	CGTWDASL SAGLF	2092	TGCGGAACATGGGATGCCAGCCTGAGTGCTGGGCT TTCTTC
1269	CGTWDASL SAEVF	2093	TGTGGCACATGGGATGCCAGCCTGAGTGCTGAAGT CTTC
1270	CGTWDASL SADFWVF	2094	TGCGGAACATGGGATGCCAGCCTGAGTGCTGACTT TTGGGTGTTC
1271	CGTWDASL RVFF	2095	TGCGGAACATGGGATGCCAGCCTGAGAGTCTTCTT C
1272	CGTWDASL RAVVL	2096	TGCGGAACATGGGATGCCAGTCTGAGGGCTGTGGT ACTCTTC
1273	CGTWDASL NIWVF	2097	TGCGGAACATGGGATGCCAGCCTGAATATTTGGGT TTCTTC
1274	CGTWDASL KNLVF	2098	TGCGGGACATGGGATGCCAGCCTGAAGAATCTGGT CTTC
1275	CGTWDASL GAWVF	2099	TGCGGAACATGGGATGCCAGCCTGGGTGCCTGGGT ATTCTTC
1276	CGTWDASL GAVVF	2100	TGCGGAACATGGGATGCCAGCCTGGGTGCTGTGGT CTTC
1277	CGTWDASL GAGVF	2101	TGCGGAACATGGGATGCCAGCCTGGGTGCGGGGGT CTTC

1278	CGTWDAR LSGLYVF	2102	TGCGGAACATGGGATGCTAGGCTGAGTGGCCTTTA TGTCTTC
1279	CGTWDAR LGGAVF	2103	TGTGGAACCTGGGATGCGAGACTGGGTGGTGCAGT CTTC
1280	CGTWDAN LRAGVF	2104	TGCGGAACATGGGATGCCAATCTGCGTGCTGGGGT CTTC
1281	CGTWDIHS GWVF	2105	TGCGGAACATGGGATGCTATCATAAGTGGTTGGGT GTTC
1282	CGTWDAG QSVWVF	2106	TGCGGAACATGGGATGCCGGCCAGAGTGTTCGGGT GTTC
1283	CGTWDAG LTGLYVF	2107	TGCGGCACATGGGATGCCGGGCTGACTGGCCTTTA TGTCTTC
1284	CGTWDAG LSVYVF	2108	TGCGGAACTTGGGATGCCGGTCTGAGTGTTCATGT CTTC
1285	CGTWDAG LSTGVF	2109	TGCGGGACATGGGATGCCGGCCTGAGTACTGGGGT CTTC
1286	CGTWDAG LSGDVF	2110	TGCGGAACATGGGATGCCGGCCTGAGTGGGGACGT TTTC
1287	CGTWDAG LSAGYVF	2111	TGCGGAACATGGGATGCCGGCCTGAGTGCTGGTTA TGTCTTC
1288	CGTWDAG LRVWVF	2112	TGCGGAACATGGGATGCCGGCCTGCGTGTTTGGGT GTTC
1289	CGTWDAG LREIF	2113	TGCGGAACATGGGATGCCGGCCTGAGGGAAATTTT C
1290	CGTWASSL SSWVF	2114	TGCGGAACATGGGCCAGCAGCCTGAGTTCCTGGGT GTTC
1291	CGTWAGSL SGHVF	2115	TGCGGAACATGGGCTGGCAGCCTGAGTGGTCATGT CTTC
1292	CGTWAGSL SAAWVF	2116	TGCGGAACATGGGCTGGCAGCCTGAGTGCCGCTTG GGTGTTC
1293	CGTWAGSL NVYWVF	2117	TGCGGAACATGGGCTGGCAGCCTGAATGTTTATTG GGTGTTC
1294	CGTWAGN LRPNWVF	2118	TGCGGAACATGGGCTGGCAACCTGAGACCTAATTG GGTGTTC
1295	CGTRGSLG GAVF	2119	TGCGGAACAAGGGGTAGCCTGGGTGGTGCGGTGTT C
1296	CGTRDTTL SVPVF	2120	TGCGGAACAAGGGATACCACCCTGAGTGTCCCGGT GTTC
1297	CGTRDTSL NIEIF	2121	TGCGGAACACGGGATACCAGCCTCAATATTGAAAT CTTC
1298	CGTRDTSL NDVF	2122	TGTGGAACACGGGATACCAGCCTGAATGATGTCTT C
1299	CGTRDTRL SIVVF	2123	TGCGGAACACGGGATACCCGCCTGAGTATTGTGGT TTTC
1300	CGTRDTILS AEVF	2124	TGCGGCACACGGGATACCATCCTGAGTGCTGAGGT GTTC
1301	CGTRDRSL SGWVF	2125	TGCGGAACACGGGATAGAAGCCTGAGTGGTTGGGT GTTC
1302	CGSWYYN VFLF	2126	TGCGGATCATGGTATTACAATGTCTTCCTTTTC

1303	CGSWHSSL NLVVF	2127	TGCGGATCTTGGCATAGCAGCCTCAACCTTGTCGTC TTC
1304	CGSWGSL SAPYVF	2128	TGCGGATCATGGGGTAGTGGCCTGAGTGCCCCCTTA TGTCTTC
1305	CGSWESGL GAWLF	2129	TGCGGTTCGTGGGAAAGCGGCCTGGGTGCTTGGCT GTTC
1306	CGSWDYG LLLF	2130	TGCGGATCCTGGGATTACGGCCTCCTACTCTTC
1307	CGSWDVSL TAVF	2131	TGCGGTTCATGGGATGTCAGCCTGACTGCTGTTTTTC
1308	CGSWDVSL NVGIF	2132	TGCGGATCCTGGGATGTCAGTCTCAATGTTGGCATT TTC
1309	CGSWDTTL RAWVF	2133	TGCGGATCATGGGATACCACCCTGCGTGCTTGGGT GTTC
1310	CGSWDTSP VRAWVF	2134	TGCGGCTCGTGGGATACCAGCCCTGTCCGTGCTTG GGTGTTC
1311	CGSWDTSL SVWVF	2135	TGCGGATCATGGGATACCAGCCTGAGTGTTTGGGT GTTC
1312	CGSWDTSL SAEVF	2136	TGCGGATCATGGGATACCAGCCTGAGTGCTGAGGT GTTC
1313	CGSWDTSL RAWVF	2137	TGCGGCTCGTGGGATACCAGCCTGCGTGCTTGGGT GTTC
1314	CGSWDTSL RAWAF	2138	TGCGGCTCGTGGGATACCAGCCTGCGTGCTTGGGC GTTC
1315	CGSWDTSL DARLF	2139	TGCGGATCATGGGATACCAGCCTGGATGCTAGGCT GTTC
1316	CGSWDTIL LVYVF	2140	TGCGGATCATGGGATACCATCCTGCTTGTCTATGTC TTC
1317	CGSWDRW QAAVF	2141	TGCGGATCATGGGATCGCTGGCAGGCTGCTGTCTT C
1318	CGSWDRSL SGYVF	2142	TGCGGATCATGGGATAGGAGCCTGAGTGGGTATGT CTTC
1319	CGSWDRSL SAYVF	2143	TGCGGATCATGGGATAGAAGCCTGAGTGCTTATGT CTTC
1320	CGSWDRSL SAVVF	2144	TGCGGATCATGGGATAGGAGCCTGAGTGCCGTGGT TTTC
1321	CGSWDNTL GVVLF	2145	TGCGGATCATGGGATAACACCTTGGGTGTTGTTCTC TTC
1322	CGSWDNRL STVIF	2146	TGCGGATCGTGGGATAACAGACTAAGTACTGTCAT CTTC
1323	CGSWDNRL NTVIF	2147	TGCGGAAGCTGGGATAATCGATTGAACACTGTGAT TTTC
1324	CGSWDLSP VRVLVF	2148	TGCGGTTCATGGGATCTCAGCCCTGTACGTGTCCTT GTGTTC
1325	CGSWDLSL SAVVF	2149	TGCGGATCATGGGATCTCAGCCTGAGTGCTGTCGT TTTC
1326	CGSWDKN LRAVLF	2150	TGCGGATCATGGGATAAAAACCTGCGTGCTGTGCT GTTC
1327	CGSWDISL SAGVF	2151	TGCGGCTCATGGGATATCAGCCTGAGTGCTGGGGT GTTC

1328	CGSWDIRL SAEVF	2152	TGCGGATCATGGGATATCAGACTGAGTGCAGAGGT CTTC
1329	CGSWDIKL NIGVF	2153	TGCGGATCATGGGACATCAAACCTGAATATTGGGGT ATTCTTC
1330	CGSWDFSL NYFVF	2154	TGCGGATCATGGGATTTTCAGTCTCAATTATTTTGTCTTC
1331	CGSWDASL STEVF	2155	TGCGGATCATGGGATGCCAGCCTGAGTACTGAGGT GTTC
1332	CGSWDAG LRGWVF	2156	TGCGGATCCTGGGATGCCGGCCTGCGTGGCTGGGT TTTC
1333	CGRWESSL GAVVF	2157	TGCGGAAGATGGGAGAGCAGCCTGGGTGCTGTGGT TTTC
1334	CGRWDFSL SAYVF	2158	TGCGGAAGATGGGATTTTAGTCTGAGTGCTTATGT CTTC
1335	CGQWDND LSVWVF	2159	TGCGGACAATGGGATAACGACCTGAGTGTTTGGGT GTTC
1336	CGPWHSSV TSGHVL	2160	TGCGGACCCTGGCATAGCAGCGTGACTAGTGGCCA CGTGCTC
1337	CGLWDASL SAPTWVF	2161	TGCGGATTATGGGATGCCAGCCTGAGTGCTCCTAC TTGGGTGTTC
1338	CGIWHTSL SAWVF	2162	TGTGGAATATGGCACACTAGCCTGAGTGCTTGGGT GTTC
1339	CGIWDYSL DTWVF	2163	TGCGGAATATGGGATTACAGCCTGGATACTTGGGT GTTC
1340	CGIWDTSL SAWVF	2164	TGCGGCATATGGGATACCAGCCTGAGTGCTTGGGT GTTC
1341	CGIWDTRL SVYVF	2165	TGCGGAATTTGGGATACCAGGCTGAGTGTTTATGT CTTC
1342	CGIWDTRL SVYIF	2166	TGCGGAATTTGGGATACCAGGCTGAGTGTTTATAT CTTC
1343	CGIWDTNL GYLF	2167	TGTGGAATATGGGATACGAATCTGGGTTATCTCTTC
1344	CGIWDTGL SAVVF	2168	TGCGGTATATGGGATACCGGCCTGAGTGCTGTGGT ATTCTTC
1345	CGIWDRSL SAWVF	2169	TGCGGAATATGGGATCGCAGCCTGAGTGCTTGGGT GTTT
1346	CGIRDTRL VYVF	2170	TGCGGAATTCGGGATACCAGGCTGAGTGTTTATGT CTTC
1347	CGGWSSRL GVGPVF	2171	TGCGGAGGATGGAGTAGCAGACTGGGTGTTGGCCC AGTGTTT
1348	CGGWGSG LSAWVF	2172	TGCGGAGGATGGGGTAGCGGCCTGAGTGCTTGGGT GTTC
1349	CGGWDTSL SAWVF	2173	TGCGGAGGATGGGATACCAGCCTGAGTGCTTGGGT GTTC
1350	CGGWDRG LDAWVF	2174	TGCGGAGGATGGGATAGGGGCCTGGATGCTTGGGT TTTC
1351	CGAWRNN VWVF	2175	TGCGGAGCATGGCGTAATAACGTGTGGGTGTTC
1352	CGAWNRR LNPHSHWV	2176	TGCGGAGCATGGAACAGGCGCCTGAATCCTCATTC TCATTGGGTGTTC

	F		
1353	CGAWHNK LSAVF	2177	TGCGGAGCCTGGCACAACAACTGAGCGCGGTCTT C
1354	CGAWGSSL RASVF	2178	TGCGGAGCATGGGGTAGCAGCCTGAGAGCTAGTGT CTTC
1355	CGAWGSG LSAVVF	2179	TGCGGAGCATGGGGTAGCGGCCTGAGTGCTTGGGT GTTC
1356	CGAWESSL SAPYVF	2180	TGCGGAGCATGGGAAAGTAGCCTGAGTGCCCTTA TGTCTTC
1357	CGAWESSL NVGLI	2181	TGCGGAGCATGGGAGAGCAGCCTCAATGTTGGACT GATC
1358	CGAWESGR SAGVVF	2182	TGCGGAGCATGGGAGAGCGGCCGAGTGCTGGGG TGGTGTTC
1359	CGAWDYS VSGWVF	2183	TGCGGAGCTTGGGATTACAGTGTGAGTGTTGGGT GTTC
1360	CGAWDYS LTAGVF	2184	TGCGGAGCATGGGATTACAGCCTGACTGCCGGAGT ATTC
1361	CGAWDYR LSAVLF	2185	TGCGGAGCCTGGGATTACAGACTGAGTGCCGTGCT ATTC
1362	CGAWDVR LDVGVF	2186	TGCGGAGCGTGGGATGTTCTGTCTGGATGTTGGGGT GTTC
1363	CGAWDTY SYVF	2187	TGCGGAGCATGGGATACCTACAGTTATGTCTTC
1364	CGAWDTTL SGVVF	2188	TGCGGAGCATGGGATACGACCCTGAGTGGTGTGGT ATTC
1365	CGAWDTTL SAVIF	2189	TGCGGAGCGTGGGATACTACCCTGAGTGCTGTGAT ATTC
1366	CGAWDTS QGASYVF	2190	TGCGGCGCATGGGATACCAGCCAGGTGCGTCTTA TGTCTTT
1367	CGAWDTSP VRAGVF	2191	TGCGGAGCATGGGATACCAGCCCTGTACGTGCTGG GGTGTTC
1368	CGAWDTSL WLF	2192	TGCGGAGCATGGGATACCAGCCTGTGGCTTTTC
1369	CGAWDTSL TVYVF	2193	TGCGGAGCATGGGATACCAGCCTGACTGTTTATGT CTTC
1370	CGAWDTSL TAGVF	2194	TGCGGAGCATGGGACACCAGTCTGACTGCTGGGGT GTTC
1371	CGAWDTSL STVVF	2195	TGCGGAGCTTGGGATACCAGCCTGAGTACTGTGGT TTTC
1372	CGAWDTSL SSRYIF	2196	TGCGGAGCATGGGATACCAGCCTGAGTTCTAGATA CATATTC
1373	CGAWDTSL SGYVF	2197	TGCGGAGCATGGGATACCAGCCTGAGTGGTTATGT CTTC
1374	CGAWDTSL SGWVF	2198	TGCGGAGCCTGGGATACCAGCCTGAGTGGCTGGGT GTTC
1375	CGAWDTSL SGVLF	2199	TGCGGAGCATGGGATACCAGTCTGAGTGGTGTGCT ATTC
1376	CGAWDTSL SGLVF	2200	TGCGGAGCTTGGGATACCAGCTTGAGTGGTCTTGT TTTC
1377	CGAWDTSL	2201	TGCGGAGCTTGGGATACCAGCTTGAGTGGTTTTGTT

	SGFVF		TTC
1378	CGAWDTSL SGEVF	2202	TGCGGAGCATGGGATACCAGCCTGAGTGGTGAGGT CTTT
1379	CGAWDTSL SDFVF	2203	TGCGGAGCTTGGGATACCAGCTTGAGTGATTTTGTT TTC
1380	CGAWDTSL RTAIF	2204	TGCGGAGCATGGGATACCAGCCTGCGAACTGCGAT ATTCTTC
1381	CGAWDTSL RLF	2205	TGCGGAGCATGGGATACCAGCCTGCGGCTTTTTC
1382	CGAWDTSL NVHVF	2206	TGCGGAGCATGGGATACCAGCCTGAATGTTTCATGT CTTC
1383	CGAWDTSL NKWVF	2207	TGCGGAGCATGGGATACCAGCCTCAATAAATGGGT GTTC
1384	CGAWDTR LSARLF	2208	TGCGGAGCATGGGATACCCGCCTCAGTGCGCGGCT GTTC
1385	CGAWDTR LRGFIF	2209	TGCGGAGCATGGGATACCAGACTGAGGGGTTTTAT TTTC
1386	CGAWDTN LGNVLL	2210	TGCGGAGCATGGGATACTAATTTGGGGAATGTTCT CCTC
1387	CGAWDTN LGKWVF	2211	TGCGGGGCATGGGATACCAACCTGGGTAAATGGGT TTTC
1388	CGAWDTG LEWYVF	2212	TGCGGAGCATGGGATACCGGCCTTGAGTGGTATGT TTTT
1389	CGAWDRTS GLWLF	2213	TGCGGAGCATGGGATAGGACTTCTGGATTGTGGCT TTTC
1390	CGAWDRSL VAGLF	2214	TGCGGAGCGTGGGATCGTAGCCTGGTTGCTGGACT CTTC
1391	CGAWDRSL TVYVF	2215	TGCGGAGCGTGGGATAGAAGCCTGACTGTTTATGT CTTC
1392	CGAWDRSL SGYVF	2216	TGCGGAGCATGGGATAGAAGCCTGAGTGGTTATGT CTTC
1393	CGAWDRSL SAYVF	2217	TGCGGAGCATGGGATAGAAGCCTGAGTGCTTATGT CTTC
1394	CGAWDRSL SAVVF	2218	TGCGGAGCATGGGATAGAAGCCTGAGTGCGGTGGT ATTCTTC
1395	CGAWDRSL SAGVF	2219	TGCGGAGCATGGGATCGCAGCCTGAGTGCTGGGGT TTTC
1396	CGAWDRSL RIVVF	2220	TGCGGAGCGTGGGATCGCAGCCTGCGTATTGTGGT ATTCTTC
1397	CGAWDRSL RAYVF	2221	TGCGGAGCATGGGATAGAAGTCTGAGGGCTTACGT CTTC
1398	CGAWDRSL NVWLF	2222	TGCGGAGCATGGGATAGAAGTCTGAATGTTTGGCT GTTC
1399	CGAWDRG LNVGWLF	2223	TGCGGGCGCCTGGGATAGGGGCCTGAATGTTCGGTTG GCTTTTC
1400	CGAWDNR LSILAF	2224	TGCGGGCGCATGGGATAATAGACTGAGTATTTTGGC CTTC
1401	CGAWDND LTAYVF	2225	TGCGGAGCTTGGGATAATGACCTGACAGCTTATGT CTTC
1402	CGAWDFSL	2226	TGCGGGGCATGGGATTTTCAGCCTGACTCCTCTCTTC

	TPLF		
1403	CGAWDDY RGVSIYVF	2227	TGCGGAGCCTGGGATGACTATCGGGGTGTGAGTAT TTATGTCTTC
1404	CGAWDDR PSSAVVF	2228	TGTGGAGCATGGGATGACCGGCCTTCGAGTGCCGT GGTTTTTC
1405	CGAWDDR LTVVVF	2229	TGCGGAGCATGGGATGACAGACTGACTGTCGTTGT TTTC
1406	CGAWDDR LGAVF	2230	TGCGGAGCGTGGGATGACAGGCTGGGTGCTGTGTT C
1407	CGAWDAS LNPGRF	2231	TGCGGAGCGTGGGATGCCAGCCTGAATCCTGGCCG GGCATTC
1408	CGAWDAG LREIF	2232	TGCGGAGCATGGGATGCCGGCCTGAGGGAAATTTT C
1409	CGAWAGSP SPWVF	2233	TGCGGAGCTTGGGCTGGCAGTCCGAGTCCTTGGGT TTTC
1410	CGAFDTTL SAGVF	2234	TGCGGAGCATTTCGACACCACCCTGAGTGCTGGCGT TTTC
1411	CETWESSL SVGVF	2235	TGCGAAACATGGGAGAGCAGCCTGAGTGTTGGGGT CTTC
1412	CETWESSL RVWVF	2236	TGCGAAACATGGGAAAGCAGCCTGAGGGTTTGGGT GTTC
1413	CETWDTSL SGGVF	2237	TGCGAAACGTGGGATACCAGCCTGAGTGGTGGGGT GTTC
1414	CETWDTSL SDFYVF	2238	TGCGAAACATGGGATACCAGCCTGAGTGACTTTTA TGTCTTC
1415	CETWDTSL SALF	2239	TGCGAAACATGGGATACCAGCCTGAGTGCCCTCTT C
1416	CETWDTSL RAEVF	2240	TGCGAAACATGGGATACCAGCCTGCGTGCTGAAGT CTTC
1417	CETWDTSL NVVVF	2241	TGCGAAACATGGGATACCAGCCTGAATGTTGTGGT ATTC
1418	CETWDTSL GAVVF	2242	TGCGAAACATGGGATACCAGCCTGGGTGCCGTGGT GTTC
1419	CETWDRSL SGVVF	2243	TGCGAAACATGGGATAGAAGCCTGAGTGGTGTGGT ATTC
1420	CETWDRSL SAWVF	2244	TGCGAAACATGGGATAGGAGCCTGAGTGCTTGGGT GTTT
1421	CETWDRSL SAVVF	2245	TGCGAAACATGGGATCGCAGCCTGAGTGCTGTGGT CTTC
1422	CETWDRGL SVVVF	2246	TGCGAGACGTGGGATAGAGGCCTGAGTGTTGTGGT TTTC
1423	CETWDRGL SAVVF	2247	TGCGAAACATGGGATAGGGGCCTGAGTGCACTGGT ATTC
1424	CETWDHTL SVVIF	2248	TGCGAAACATGGGATCACACCCTGAGTGTTGTGAT ATTC
1425	CETWDASL TVVLF	2249	TGCGAAACATGGGATGCCAGCCTGACTGTTGTGTT ATTC
1426	CETWDASL SAGVF	2250	TGCGAAACATGGGATGCCAGCCTGAGTGCTGGGGT GTTC
1427	CETWDAGL	2251	TGCGAAACGTGGGATGCCGGCCTGAGTGAGGTGGT

	SEVVF		GTTC
1428	CETFDTSLS VVVF	2252	TGCGAAACATTTGATACCAGCCTGAGTGTGTAGT CTTC
1429	CETFDTSLS NIVVF	2253	TGCGAAACATTTGATACCAGCCTAAATATTGTAGT CTTT
1430	CESWDRSR IGVVF	2254	TGCGAATCATGGGATAGAAGCCGGATTGGTGTGGT CTTC
1431	CESWDRSL SARVY	2255	TGCGAAAGTTGGGACAGGAGTCTGAGTGCCCGGGT GTAC
1432	CESWDRSL RAVVF	2256	TGCGAATCCTGGGATAGGAGCCTGCGTGCCGTGGT CTTC
1433	CESWDRSL IVVF	2257	TGCGAATCTTGGGATCGTAGTTTGATTGTGGTGTTC
1434	CESWDNNL NEVVF	2258	TGCGAAAGTTGGGATAACAATTTAAATGAGGTGGT TTTC
1435	CEIWESSPS ADDLVF	2259	TGCGAAATATGGGAGAGCAGCCCGAGTGCTGACG ATTTGGTGTTC
1436	CEAWDTSLS SGAVF	2260	TGCGAAGCATGGGATACCAGCCTGAGTGGTGCGGT GTTC
1437	CEAWDTSLS SAGVF	2261	TGCGAAGCATGGGATACCAGCCTGAGTGCCGGGGT GTTC
1438	CEAWDTSLS GGGVF	2262	TGCGAAGCATGGGATACCAGCCTGGGTGGTGGGGT GTTC
1439	CEAWDRSL TGS LF	2263	TGCGAAGCATGGGATCGCAGCCTGACTGGTAGCCT GTTC
1440	CEAWDRG LSAVVF	2264	TGCGAAGCGTGGGATAGGGGCCTGAGTGCA GTGGT ATTC
1441	CEAWDNIL STVVF	2265	TGCGAAGCCTGGGATAACATCCTGAGTACTGTGGT GTTC
1442	CEAWDISL SAGVF	2266	TGCGAAGCATGGGACATCAGCCTGAGTGCTGGGGT GTTC
1443	CEAWDAD LSGAVF	2267	TGCGAAGCATGGGATGCCGACCTGAGTGGTGCGGT GTTC
1444	CATWTGSF RTGHYVF	2268	TGCGCAACATGGACTGGTAGTTTCAGAACTGGCCA TTATGTCTTC
1445	CATWSSSP RGWVF	2269	TGCGCAACATGGAGTAGCAGTCCCAGGGGGTGGGT GTTC
1446	CATWHYSL SAGRVF	2270	TGCGCAACATGGCATTACAGCCTGAGTGCTGGCCG AGTGTTC
1447	CATWHTSL SIVQF	2271	TGCGCAACATGGCATACCAGCCTGAGTATTGTGCA GTTC
1448	CATWHSTL SADVLF	2272	TGCGCAACATGGCATAGCACCTGAGTGCTGATGT GCTTTTC
1449	CATWHSSL SAGR LF	2273	TGCGCAACATGGCATAGCAGCCTGAGTGCTGGCCG ACTCTTC
1450	CATWHIAR SAWVF	2274	TGCGCAACATGGCATATCGCTCGGAGTGCTGGGT GTTC
1451	CATWGSSQ SAVVF	2275	TGCGCAACATGGGGTAGTAGTCAGAGTGCCGTGGT ATTC
1452	CATWGSSL	2276	TGCGCAACATGGGGTAGCAGCCTGAGTGCTGGGGG

	SAGGVF		TGTTTTTC
1453	CATWEYSL SVVLF	2277	TGTGCAACATGGGAATACAGCCTGAGTGTTGTGCT GTTC
1454	CATWETTR RASVVF	2278	TGCGCAACATGGGAGACCACCCGACGTGCCTCTTT TGTCTTC
1455	CATWETSL NVYVF	2279	TGCGCAACATGGGAGACCAGCCTGAATGTTTATGT CTTC
1456	CATWETSL NVVVF	2280	TGCGCAACATGGGAAACTAGCCTGAATGTTGTGGT CTTC
1457	CATWETSL NLYVF	2281	TGCGCAACATGGGAGACCAGCCTGAATCTTTATGT CTTC
1458	CATWETGL SAGEVF	2282	TGCGCAACATGGGAGACTGGCCTAAGTGCTGGAGA GGTGTTTC
1459	CATWESTL SVVVF	2283	TGCGCGACGTGGGAGAGTACCCTAAGTGTTGTGGT TTTC
1460	CATWESSL SIFVF	2284	TGCGCAACGTGGGAGAGCAGCCTGAGTATTTTTGT CTTC
1461	CATWESSL NTFYVF	2285	TGCGCAACATGGGAAAGCAGCCTCAACACTTTTTA TGTCTTC
1462	CATWESRV DTRGLLF	2286	TGCGCAACATGGGAGAGTAGGGTGGATACTCGAG GGTTGTTATTC
1463	CATWESGL SGAGVF	2287	TGCGCAACATGGGAGAGCGGCCTGAGTGGTGC GG GGGTGTTTC
1464	CATWEGSL NTFYVF	2288	TGCGCAACATGGGAAGGCAGCCTCAACACTTTTTA TGTCTTC
1465	CATWDYSL SAVVF	2289	TGCGCAACTTGGGATTATAGCCTGAGTGCTGTGGT GTTC
1466	CATWDYR LSIVVF	2290	TGCGCAACATGGGATTACAGACTGAGTATTGTGGT ATTC
1467	CATWDYN LGA AVF	2291	TGCGCAACATGGGATTATAACCTGGGAGCTGCGGT GTTC
1468	CATWDVTL GVLHF	2292	TGCGCCACATGGGATGT CACCCTGGGTGTCTTGCA TTTC
1469	CATWD TTL SVWVF	2293	TGCGCAACATGGGATACAACACTGAGTGTCTGGGT CTTC
1470	CATWD TTL SVVLF	2294	TGCGCAACATGGGATACCACCCTGAGTGTAGTACT TTTC
1471	CATWD TTL SVEVF	2295	TGCGCAACATGGGATACCACCCTGAGTGTTGAGGT CTTC
1472	CATWD TSP SLSGFWVF	2296	TGCGCAACATGGGATACCAGCCCCAGCCTGAGTGG TTTTTGGGTGTTTC
1473	CATWD TSL TGVVF	2297	TGCGCAACATGGGATACCAGCCTGACTGGTGTGGT ATTC
1474	CATWD TSL TGAVF	2298	TGCGCAACATGGGATACCAGCCTGACTGGTGC GG GTTC
1475	CATWD TSL TAWVF	2299	TGCGCAACATGGGATACCAGCCTGACTGCCTGGGT ATTC
1476	CATWD TSL TAVVF	2300	TGCGCAACATGGGATACCAGCCTGACTGCTGTGGT TTTC
1477	CATWD TSL	2301	TGCGCAACATGGGATACTAGCCTGACTGCTAAGGT

	TAKVF		GTTC
1478	CATWDTSL SVVVF	2302	TGCGCAACATGGGACACCAGCCTGAGTGTTGTGGT TTTC
1479	CATWDTSL SVGVF	2303	TGCGCTACTTGGGATACCAGCCTGAGTGTTGGGGT ATTT
1480	CATWDTSL SSWVF	2304	TGCGCAACATGGGATACCAGCCTGAGTTCTTGGGT GTTC
1481	CATWDTSL SGGVF	2305	TGCGCAACATGGGATACCAGCCTGAGTGGTGGGGT ACTC
1482	CATWDTSL SGGVF	2306	TGCGCAACATGGGATACCAGCCTGAGTGGTGGGGT GTTC
1483	CATWDTSL SGGRVF	2307	TGCGCAACATGGGATACCAGCCTGAGTGGTGGCCG AGTGTTC
1484	CATWDTSL SGDRVF	2308	TGCGCAACATGGGATACCAGCCTGAGTGGTGACCG AGTGTTC
1485	CATWDTSL SEGVF	2309	TGCGCAACGTGGGATACTAGCCTGAGTGAAGGGGT GTTC
1486	CATWDTSL SAVVL	2310	TGCGCAACCTGGGATACCAGCCTGAGTGCCGTGGT GCTC
1487	CATWDTSL SAVF	2311	TGCGCAACATGGGATACCAGCCTGAGTGCTGTCTT C
1488	CATWDTSL SARVF	2312	TGCGCGACATGGGATACCAGCCTGAGTGCTCGGGT GTTC
1489	CATWDTSL SALF	2313	TGCGCAACATGGGATACCAGCCTGAGTGCCTTATT C
1490	CATWDTSL SAHVF	2314	TGCGCAACATGGGATACCAGCCTGAGTGCTCATGT CTTC
1491	CATWDTSL SAGRVF	2315	TGCGCAACATGGGATACCAGCCTGAGTGCTGGCCG GGTGTTC
1492	CATWDTSL SAEVF	2316	TGCGCAACATGGGATACCAGCCTGAGTGCGGAGGT CTTC
1493	CATWDTSL SADAGGV F	2317	TGCGCAACATGGGATACCAGCCTGAGTGCTGATGC TGGTGGGGGGGTCTTC
1494	CATWDTSL RVVVF	2318	TGCGCAACATGGGATACCAGCCTGCGTGTCGTGGT ATTC
1495	CATWDTSL RGVF	2319	TGCGCAACATGGGATACCAGCCTGAGAGGGGTGTT C
1496	CATWDTSL PAWVF	2320	TGCGCAACATGGGATACCAGCCTGCCTGCGTGGGT GTTC
1497	CATWDTSL NVGVF	2321	TGTGCAACATGGGATACCAGCCTGAATGTTGGGGT ATTC
1498	CATWDTSL GIVLF	2322	TGCGCAACATGGGATACCAGCCTGGGTATTGTGTT ATTT
1499	CATWDTSL GARVVF	2323	TGCGCAACATGGGACACCAGCCTGGGTGCGCGTGT GGTCTTC
1500	CATWDTSL GALF	2324	TGTGCAACGTGGGATACCAGTCTAGGTGCCTTGTT C
1501	CATWDTSL ATGLF	2325	TGCGCAACATGGGATACCAGCCTGGCGACTGGACT GTTC

1502	CATWDTSL AAWVF	2326	TGCGCAACATGGGATACCAGCCTGGCTGCCTGGGT ATTC
1503	CATWDTRL SAVVF	2327	TGCGCAACCTGGGATACCAGGCTGAGTGCTGTGGT CTTC
1504	CATWDTRL SAGVF	2328	TGCGCAACATGGGATACCAGGCTGAGTGCTGGGGT GTTC
1505	CATWDTRL LITVF	2329	TGTGCAACGTGGGACACACGTCTACTTATTACGGT TTTC
1506	CATWDTRL SVELF	2330	TGCGCAACATGGGACACCCTCCTGAGTGTTGAACT CTTC
1507	CATWDTG RNPHVVF	2331	TGCGCAACATGGGATACTGGCCGCAATCCTCATGT GGTCTTC
1508	CATWDTGL SSVLF	2332	TGCGCAACATGGGATACCGGCCTGTCTTCGGTGTT GTTC
1509	CATWDTGL SAVF	2333	TGCGCAACGTGGGATACCGGCCTGAGTGCGGTTTT C
1510	CATWDRTL SIGVF	2334	TGCGCTACGTGGGATAGGACCCTGAGTATTGGAGT CTTC
1511	CATWDRSV TAVLF	2335	TGCGCAACGTGGGATCGCAGTGTGACTGCTGTGCT CTTC
1512	CATWDRSL SGVVF	2336	TGCGCAACCTGGGATAGGAGCCTGAGTGGTGTGGT GTTC
1513	CATWDRSL SAVVF	2337	TGCGCAACATGGGATAGAAGCCTGAGTGCTGTGGT CTTC
1514	CATWDRSL SAVPWVF	2338	TGCGCAACATGGGATAGAAGCCTGAGTGCTGTTC TTGGGTGTTC
1515	CATWDRSL SAGVF	2339	TGCGCAACATGGGATCGCAGCCTGAGTGCTGGGGT GTTC
1516	CATWDRSL RAGVF	2340	TGCGCAACGTGGGATAGGAGCCTGCGTGCTGGGGT GTTC
1517	CATWDRSL NVYVL	2341	TGCGCAACATGGGATCGCAGTCTGAATGTTTATGT CCTC
1518	CATWDRIL SAEVF	2342	TGCGCAACGTGGGATCGCATCCTGAGCGCTGAGGT GTTC
1519	CATWDRG LSTGVF	2343	TGCGCAACGTGGGATAGAGGCCTGAGTACTGGGGT GTTC
1520	CATWDNY LGAADF	2344	TGCGCAACATGGGATAACTACCTGGGTGCTGCCGT GTTC
1521	CATWDNTP SNIVVF	2345	TGCGCAACATGGGATAACACGCCTTCGAATATTGT GGTATTC
1522	CATWDNTL SVVVF	2346	TGCGCAACATGGGATAATACACTGAGTGTGTGGGT CTTC
1523	CATWDNTL SVNWVF	2347	TGCGCAACATGGGATAACACCCTGAGTGTC AATTG GGTGTTC
1524	CATWDNTL NVFYVF	2348	TGCGCAACCTGGGATAACACACTGAATGTCTTTTA TGTTTTTC
1525	CATWDNR LSSVVF	2349	TGTGCGACATGGGATAATCGGCTCAGTTCTGTGGT CTTC
1526	CATWDNR LSAGVL	2350	TGCGCAACATGGGATAACCGCCTGAGTGCTGGGGT GCTC

1527	CATWDNR LSAGVF	2351	TGCGCAACGTGGGATAACAGGCTGAGTGCTGGGGT GTTC
1528	CATWDNR DWVF	2352	TGCGCAACATGGGATAACAGGGATTGGGTCTTC
1529	CATWDNN LGAGVF	2353	TGCGCAACATGGGATAACAACCTGGGTGCTGGGGT GTTC
1530	CATWDNK LTSGVF	2354	TGCGCAACATGGGATAACAAGCTGACTTCTGGGGT CTTC
1531	CATWDNIL SAWVF	2355	TGCGCAACATGGGATAACATCCTGAGTGCCTGGGT GTTT
1532	CATWDNDI HSGLF	2356	TGCGCAACCTGGGACAACGATATACATTCTGGGCT GTTC
1533	CATWDLSL SALF	2357	TGCGCAACTTGGGATCTCAGCCTGAGTGCCCTGTTC
1534	CATWDITL SAEVF	2358	TGCGCAACATGGGATATCACCTGAGTGCTGAGGT GTTC
1535	CATWDISP SAGGVF	2359	TGCGCAACGTGGGATATCAGCCCGAGTGCTGGCGG GGTGTTC
1536	CATWDISL STGRAVF	2360	TGCGCAACATGGGATATCAGTCTAAGTACTGGCCG GGCTGTGTTC
1537	CATWDISL SQVF	2361	TGCGCAACATGGGATATCAGTCTGAGTCAGGTATT C
1538	CATWDIRL SSGVF	2362	TGCGCAACATGGGATATCAGGCTGAGTAGTGAGGT GTTC
1539	CATWDIGP SAGGVF	2363	TGCGCAACGTGGGATATCGGCCCGAGTGCTGGCGG GGTGTTC
1540	CATWDHSR AGVLF	2364	TGCGCAACATGGGATCACAGCCGGGCTGGTGTGCT ATTCTTC
1541	CATWDHSP SVGEVF	2365	TGCGCAACATGGGATCACAGTCCGAGTGTTGGAGA AGTCTTC
1542	CATWDHSL RVGVF	2366	TGCGCAACATGGGATCACAGCCTGCGTGTTGGGGT GTTC
1543	CATWDHSL NIGVF	2367	TGCGCAACATGGGATCACAGCCTGAACATTGGGGT GTTC
1544	CATWDHSL GLWAF	2368	TGCGCAACATGGGATCACAGCCTGGGTCTTTGGGC ATTCTTC
1545	CATWDHN LRLVF	2369	TGCGCCACATGGGATCACAATCTGCGTCTTGTTTTT
1546	CATWDHIL ASGVF	2370	TGCGCGACTTGGGATCACATCCTGGCTTCTGGGGT GTTC
1547	CATWDFSL SVWVF	2371	TGCGCAACATGGGATTTTCAGCCTGAGTGTTTGGGT GTTC
1548	CATWDFSL SAWVF	2372	TGCGCAACATGGGATTTTCAGCCTGAGTGCTTGGGT GTTC
1549	CATWDDTL TAGVF	2373	TGCGCAACATGGGATGACACCCTCACTGCTGGTGT GTTC
1550	CATWDDR LSAVLF	2374	TGCGCAACATGGGACGACAGGCTGAGTGCTGTGCT TTTC
1551	CATWDDR LDAAVF	2375	TGCGCAACATGGGATGACAGGCTGGATGCTGCGGT GTTC

1552	CATWDATL NTGVF	2376	TGCGCAACATGGGATGCGACCCTGAATACTGGGGT GTTC
1553	CATWDASL SVWLL	2377	TGCGCAACATGGGATGCCAGCCTGAGTGTTTGGCT GCTC
1554	CATWDASL SGGVF	2378	TGCGCGACATGGGATGCCAGCCTGAGTGGTGGGGT GTTC
1555	CATRD TTL SAVLF	2379	TGCGCAACACGGGATACCACCCTCAGCGCCGTTCT GTTC
1556	CATLGSSLS LWVF	2380	TGCGCTACATTGGGTAGTAGCCTGAGTCTCTGGGT GTTC
1557	CATIETSLP AWVF	2381	TGCGCAACAATCGAACTAGCCTGCCTGCCTGGGT ATTCTTC
1558	CATGDRSL TVEVF	2382	TGCGCAACAGGGGACAGAAGCCTGACTGTTGAGGT ATTCTTC
1559	CATGDLGL TIVF	2383	TGCGCTACAGGGGATCTCGGCCTGACCATAGTCTT C
1560	CASWDYR GRSGWVF	2384	TGCGCATCATGGGATTACAGGGGGAGATCTGGTTG GGTGTTC
1561	CASWD TTL NVGVF	2385	TGCGCATCATGGGATACCACCCTGAATGTTGGGGT GTTC
1562	CASWD TTL GFVLF	2386	TGCGCTTCATGGGATACCACCCTGGGTTTTGTGTTA TTC
1563	CASWD TSL SGGYVF	2387	TGCGCATCATGGGATACCAGCCTGAGTGGTGGTTA TGTCTTC
1564	CASWD TSL RAGVF	2388	TGCGCATCATGGGATACCAGCCTCCGTGCTGGGGT GTTC
1565	CASWD TSL GAGVF	2389	TGCGCATCATGGGATACCAGCCTGGGTGCTGGGGT GTTC
1566	CASWDRGL SAVVF	2390	TGCGCATCATGGGACAGAGGCCTGAGTGCAGTGGT GTTC
1567	CASWDNV LRGVVF	2391	TGTGCTAGTTGGGATAACGTCCTGCGTGGTGTGGT ATTCTTC
1568	CASWDNRL TAVVF	2392	TGCGCGTCATGGGATAACAGGCTGACTGCCGTGGT TTTC
1569	CASWDASL SVAFF	2393	TGCGCATCATGGGATGCAAGCCTGTCCGTGCTTTTC
1570	CASWDAG LSSYVF	2394	TGCGCTTCGTGGGATGCCGGCCTGAGTTCTTATGTC TTC
1571	CASGDTSL SGVIF	2395	TGCGCATCCGGGGATACCAGCCTGAGTGGTGTGAT ATTCTTC
1572	CARWHTSL SIWVF	2396	TGCGCAAGATGGCATAACGAGCCTAAGTATTTGGGT CTTC
1573	CAIWD TGL SPGQVAF	2397	TGCGCAATATGGGATACCGGCCTGAGTCCTGGCCA AGTTGCCTTC
1574	CAAWHSG LGLPVF	2398	TGCGCAGCATGGCATAGCGGCCTGGGTCTCCCGGT CTTC
1575	CAAWDYS LSAGVF	2399	TGCGCAGCATGGGATTACAGCCTGAGTGCTGGGGT GTTC
1576	CAAWD TTL RVRLF	2400	TGCGCAGCCTGGGATACTACCCTGCGTGTTAGGCT GTTC

1577	CAAWDTSL TAWVF	2401	TGCGCAGCATGGGATACCAGCCTGACTGCCTGGGT TTTC
1578	CAAWDTSL SGGVF	2402	TGCGCAGCATGGGATACCAGCTTGAGTGGTGGGGT GTTC
1579	CAAWDTSL SGEAVF	2403	TGCGCAGCATGGGATACCAGCCTGAGTGGCGAGGC TGTGTTC
1580	CAAWDTSL SGAVF	2404	TGCGCAGCATGGGATACCAGCTTGAGTGGTGCGGT GTTC
1581	CAAWDTSL SAWVF	2405	TGCGCAGCATGGGATACCAGCCTGAGTGCCTGGGT GTTC
1582	CAAWDTSL SAGVF	2406	TGCGCAGCATGGGATACCAGCCTGAGTGCTGGGGT ATTTC
1583	CAAWDTSL DTYVF	2407	TGCGCAGCATGGGATACCAGCCTGGATACTTATGT CTTC
1584	CAAWDTR LSGVLF	2408	TGCGCTGCATGGGATACCCGTCTGAGTGGTGTGTT ATTTC
1585	CAAWDTR LSAGVF	2409	TGCGCAGCATGGGATACCAGGCTGAGTGCTGGGGT GTTC
1586	CAAWDRSL STGVF	2410	TGCGCAGCATGGGATCGCAGTCTGAGTACTGGAGT TTTC
1587	CAAWDIRR SVLF	2411	TGCGCAGCGTGGGATATCCGCCGGTCTGTCCTTTTC
1588	CAAWDHT QRLSF	2412	TGCGCTGCGTGGGATCACACTCAGCGTCTTTCCTTC
1589	CAAWDHS LSAGQVF	2413	TGCGCAGCATGGGATCACAGCCTGAGTGCTGGCCA GGTGTTC
1590	CAAVDTGL KEWVF	2414	TGCGCAGCAGTCGATACTGGTCTGAAAGAATGGGT GTTC

[00210] The CDRs were prescreened to contain no amino acid liabilities, cryptic splice sites or nucleotide restriction sites. The CDR variation was observed in at least two individuals and comprises the near-germline space of single, double and triple mutations. The order of assembly is seen in **FIG. 21C**.

[00211] The VH domains that were designed include IGHV1-69 and IGHV3-30. Each of two heavy chain VH domains are assembled with their respective invariant 4 framework elements (FW1, FW2, FW3, FW4) and variable 3 CDR (H1, H2, H3) elements. For IGHV1-69, 417 variants were designed for H1 and 258 variants were designed for H2. For IGHV3-30, 535 variants were designed for H1 and 165 variants were designed for H2. For the CDR H3, the same cassette was used in both IGHV1-69 and IGHV3-30 since both designed use an identical FW4, and because the edge of FW3 is also identical for both IGHV1-69 and IGHV3-30. The CDR H3 comprises an N-terminus and C-terminus element that are combinatorially joined to a central middle element to generate 1×10^{10} diversity. The N-terminal and middle element overlap with a “GGG” glycine codon. The middle and C-terminal element overlap with a

“GGT” glycine codon. The CDR H3 comprises 5 subpools that were assembled separately. The various N-terminus and C-terminus elements comprise sequences as seen in **Table 14**.

Table 14. Sequences for N-terminus and C-terminus elements

Element	SEQ ID NO	Sequence
Stem A	2415	CARDLRELECEEW XXX SRGPCVDPRGVAGSFDVW
Stem B	2416	CARDMYDYDF XXX EVVPADDAFDIW
Stem C	2417	CARDGRGSLPRPKGGP XXX YDSSEDSGGAFDIW
Stem D	2418	CARANQHF XXX GYHYYGMDVW
Stem E	2419	CAKHMSMQ XXX RADLVGDAFDVW

[00212] Example 12. Enrichment for GPCR GLP1R Binding Proteins

[00213] Antibodies having CDR-H3 regions with a variant fragments of GPCR binding protein were generated by methods described herein were panned using cell-based methods to identified variants which are enriched for binding to particular GPCRs, as described in Example 10.

[00214] Variants of the GLP C-terminus peptide were identified (listed in **Table 15**) that when embedded in the CDR-H3 region of an antibody, were repeatedly and selectively enriched for binding to GPCR GLP1R.

Table 15. Sequences of GLP1 embedded in CDR-H3

SEQ ID NO	Sequence
2420	CAKHMSMQEGAVTGEGQAAKEFIAWLKGRVRADLVGDAFDVW
2421	CARDGRGSLPRPKGGPQTVGEGQAAKEFIAWLKGGTYDSEDSGGAFDIW
2422	CAKHMSMQDYLVIGEGQAAKEFIAWLKGGPARADLVGDAFDVW
2423	CAKHMSMQEGAVTGEGQDAKEFIAWLKGRVRADLVGDAFDVW
2424	WAKHMSMQEGAVTGEGQAAKEFIAWLKGRVRADLVGDAFDVW
2425	CARDGRGSLPRPKGGPQTVGEGQAAKEFIAWLKGRVRADLVGDAFDVW
2426	CARANQHFEQEGTFTSDVSSYLEGQAAKEFIAWLKGGIRGYHYYGMDVW
2427	CARANQHFTLHGEQAAKEFIAWLKGRGQIDIGYHYYGMDVW
2428	CARANQHFLGAGVSSYLEGQAAKEFIAWLKGDTTGYHYYGMDVW
2429	CARANQHFLDKGTFTSDVSSYLEGQAAKEFIAWLKGIYPGYHYYGMDVW
2430	CARANQHFGTSLAGEGQAAKEFIAWLKGGSQYDSEDSGGAFDIW
2431	CARANQHFLHAGGEGQAAKEFIAWLKGGSGTYGYHYYGMDVW
2432	CARANQHFGGKGEGQAAKEFIAWLKGGSGAGYHYYGMDVW
2433	CAKQMSMQEGAVTGEGQAAKEFIAWLKGRVRADLVGDAFDVW
2434	CAKHMSMQEGAVTGEGQAAKEFIAWLKGGPARADLVGDAFDVW

2435	CAKHMSMQEGAVTGEGQAAKEFIAWLVKGGLTYDSSSEDSGGAFDIW
2436	CAKHMSMQDYLVIGEGQAAKEFIAWLVKGRVRADLVGDADFVW

[00215] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein may be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. An antibody, wherein the antibody comprises a CDR-H3 comprising a sequence of any one of SEQ ID NOS: 2420 to 2436.
2. An antibody, wherein the antibody comprises a CDR-H3 comprising a sequence of any one of SEQ ID NOS: 2420 to 2436; and wherein the antibody is a monoclonal antibody, a polyclonal antibody, a bi-specific antibody, a multispecific antibody, a grafted antibody, a human antibody, a humanized antibody, a synthetic antibody, a chimeric antibody, a camelized antibody, a single-chain Fvs (scFv), a single chain antibody, a Fab fragment, a F(ab')₂ fragment, a Fd fragment, a Fv fragment, a single-domain antibody, an isolated complementarity determining region (CDR), a diabody, a fragment comprised of only a single monomeric variable domain, disulfide-linked Fvs (sdFv), an intrabody, an anti-idiotypic (anti-Id) antibody, or ab antigen-binding fragments thereof.
3. A method of inhibiting GLP1R activity, comprising administering the antibody of claim 1 or 2.
4. A method for treatment of a metabolic disorder, comprising administering to a subject in need thereof the antibody of claim 1 or 2.
5. The method of claim 4, wherein the metabolic disorder is Type II diabetes or obesity.
6. A nucleic acid library comprising a plurality of nucleic acids, wherein each nucleic acid encodes for a sequence that when translated encodes for an immunoglobulin scaffold, wherein the immunoglobulin scaffold comprises a CDR-H3 loop that comprises a GPCR binding domain, and wherein each nucleic acid comprises a sequence encoding for a sequence variant of the GPCR binding domain.
7. The nucleic acid library of claim 6, wherein a length of the CDR-H3 loop is about 20 to about 80 amino acids.
8. The nucleic acid library of claim 6, wherein a length of the CDR-H3 loop is about 80 to about 230 base pairs.
9. The nucleic acid library of claim 6, wherein the immunoglobulin scaffold further comprises one or more domains selected from variable domain, light chain (VL), variable domain, heavy chain (VH), constant domain, light chain (CL), and constant domain, heavy chain (CH).
10. The nucleic acid library of claim 9, wherein the VH domain is IGHV1-18, IGHV1-69, IGHV1-8 IGHV3-21, IGHV3-23, IGHV3-30/33m, IGHV3-28, IGHV3-74, IGHV4-39, or IGHV4-59/61.

11. The nucleic acid library of claim 9, wherein the VL domain is IGKV1-39, IGKV1-9, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV4-1, IGLV1-51, or IGLV2-14.
12. The nucleic acid library of claim 9, wherein a length of the VH domain is about 90 to about 100 amino acids.
13. The nucleic acid library of claim 9, wherein a length of the VL domain is about 90 to about 120 amino acids.
14. The nucleic acid library of claim 9, wherein a length of the VH domain is about 280 to about 300 base pairs.
15. The nucleic acid library of claim 9, wherein a length of the VL domain is about 300 to about 350 base pairs.
16. The nucleic acid library of claim 6, wherein the library comprises at least 10^5 non-identical nucleic acids.
17. The nucleic acid library of claim 6, wherein the immunoglobulin scaffold comprises a single immunoglobulin domain.
18. The nucleic acid library of claim 6, wherein the immunoglobulin scaffold comprises a peptide of at most 100 amino acids.
19. The nucleic acid library of claim 6, wherein the GPCR binding domains comprise peptidomimetic or small molecule mimetic.
20. A protein library comprising a plurality of proteins, wherein each of the proteins of the plurality of proteins comprise an immunoglobulin scaffold, wherein the immunoglobulin scaffold comprises a CDR-H3 loop that comprises a sequence variant of a GPCR binding domain.
21. The protein library of claim 20, wherein a length of the CDR-H3 loop is about 20 to about 80 amino acids.
22. The protein library of claim 20, wherein the immunoglobulin scaffold further comprises one or more domains selected from variable domain, light chain (VL), variable domain, heavy chain (VH), constant domain, light chain (CL), and constant domain, heavy chain (CH).
23. The protein library of claim 20, wherein the VH domain is IGHV1-18, IGHV1-69, IGHV1-8 IGHV3-21, IGHV3-23, IGHV3-30/33m, IGHV3-28, IGHV3-74, IGHV4-39, or IGHV4-59/61.
24. The protein library of claim 20, wherein the VL domain is IGKV1-39, IGKV1-9, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV4-1, IGLV1-51, or IGLV2-14.
25. The protein library of claim 20, wherein a length of the VH domain is about 90 to about 100 amino acids.

26. The protein library of claim 20, wherein a length of the VL domain is about 90 to about 120 amino acids.
27. The protein library of claim 20, wherein the plurality of proteins are used to generate a peptidomimetic library.
28. The protein library of claim 20, wherein the protein library comprises antibodies.
29. A protein library comprising a plurality of proteins, wherein the plurality of proteins comprises sequence encoding for different GPCR binding domains, and wherein the length of each GPCR binding domain is about 20 to about 80 amino acids.
30. The protein library of claim 29, wherein the protein library comprises peptides.
31. The protein library of claim 29, wherein the protein library comprises immunoglobulins.
32. The protein library of claim 29, wherein the protein library comprises antibodies.
33. The protein library claim 29, wherein the plurality of proteins is used to generate a peptidomimetic library.
34. A vector library comprising the nucleic acid library of any one of claims 6-19.
35. A cell library comprising the nucleic acid library of any one of claims 6-19.
36. A cell library comprising the protein library of any one of claims 20-33.