VARIANT NUCLEIC ACID LIBRARIES FOR ANTIBODY OPTIMIZATION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/810,379 filed on February 26, 2019 and U.S. Provisional Patent Application No. 62/830,296 filed on April 5, 2019, each of which is incorporated by reference in its entirety.

BACKGROUND

[0002] Antibodies possess the capability to bind with high specificity and affinity to biological targets. However, the design of therapeutic antibodies is challenging due to balancing of immunological effects with efficacy. Thus, there is a need to develop compositions and methods for the optimization of antibody properties.

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 methods, compositions, and systems for the optimization of antibodies.

[0005] Provided herein are nucleic acid libraries comprising a plurality of sequences encoding for antibodies or antibody fragments, wherein each sequence of the plurality of sequences comprises a predetermined number of mutations relative to an input sequence; the library comprises at least 5,000 variant sequences, wherein each of the at least 5,000 variant sequences is represented in an amount of no more than 50% of an amount for any other variant sequence amount in the library; and at least one sequence encodes for an antibody or antibody fragment having a higher binding affinity than the input sequence. Further provided herein are nucleic acid libraries, wherein the library comprises at least 50,000 variant sequences. Further provided herein are nucleic acid libraries, wherein the library comprises at least 100,000 variant sequences. Further provided herein are nucleic acid libraries, wherein at least some of the sequences encode for an antibody light chain. Further provided herein are nucleic acid libraries, wherein at least some of the sequences encode for an antibody heavy chain. Further provided herein are nucleic acid libraries, wherein each sequence of the plurality of sequences comprises at least one mutation in each CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are nucleic

acid libraries, wherein each sequence of the plurality of sequences comprises at least two mutations in each CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are nucleic acid libraries, wherein at least one of the mutations is present in at least two individuals. Further provided herein are nucleic acid libraries, wherein at least one of the mutations is present in at least three individuals. Further provided herein are nucleic acid libraries, wherein each sequence of the plurality of sequences comprises at least one mutation in each CDR of a heavy chain or light chain relative to a germline sequence of the input sequence.

[0006] Provided herein are antibodies, wherein the antibody comprises a CDR-H3 comprising a sequence of any one of SEQ ID NOs: 1-35. Provided herein are antibodies, wherein the antibody comprises a CDR-H3 comprising a sequence of any one of any one of SEQ ID NOs: 1-35; 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.

[0007] Provided herein are methods of inhibiting PD-1 activity, comprising administering an antibody as described herein. Provided herein are methods for treatment of a proliferative disorder, comprising administering to a subject in need thereof an antibody as described herein. Further provided herein are methods for treatment of a proliferative disorder, wherein the proliferative disorder is cancer. Further provided herein are methods for treatment of a proliferative disorder wherein the cancer is lung, head and neck squamous cell, colorectal, melanoma, liver, classical Hodgkin lymphoma, kidney, gastric, cervical, merkel cell, B-cell lymphoma, or bladder cancer.

[0008] Provided herein are nucleic acid libraries comprising a plurality of nucleic acids, wherein each nucleic acid of the plurality of nucleic acids encodes for a sequence that when translated encodes for an antibody, wherein the antibody comprises a CDR-H3 loop that comprises a PD-1 binding domain, and wherein each nucleic acid of the plurality of nucleic acids comprises a sequence encoding for a sequence variant of the PD-1 binding domain. Further provided herein are nucleic acid libraries, wherein a length of the CDR-H3 loop when translated is about 20 to about 80 amino acids. Further provided herein are nucleic acid libraries, wherein a length of the CDR-H3 loop is about 80 to about 230 base pairs. Further provided herein are nucleic acid libraries, wherein the antibody 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). Further provided herein are nucleic acid libraries, wherein a length of the VH domain is about 90 to about 100 amino acids. Further provided herein are nucleic acid libraries, wherein a length of the VL domain is about 90 to about 120 amino acids. Further provided herein are nucleic acid libraries, wherein a length of the VH domain is about 280 to about 300 base pairs. Further provided herein are nucleic acid libraries, wherein a length of the VL domain is about 300 to about 350 base pairs. Further provided herein are nucleic acid libraries, wherein the library comprises at least 10^{10} non-identical nucleic acids. Further provided herein are nucleic acid libraries, wherein the library comprises at least 10^{12} non-identical nucleic acids. Further provided herein are nucleic acid libraries, wherein the antibody comprises a single immunoglobulin domain. Further provided herein are nucleic acid libraries, wherein the antibody comprises a peptide of at most 100 amino acids. Further provided herein are nucleic acid libraries, wherein the PD-1 binding domains comprise a peptidomimetic or small molecule mimetic.

[0009] Provided herein are protein libraries comprising a plurality of proteins, wherein each protein of the plurality of proteins comprises an antibody, wherein the antibody comprises a CDR-H3 loop that comprises a sequence variant of a PD-1 binding domain. Further provided herein are protein libraries comprising a plurality of proteins, wherein a length of the CDR-H3 loop is about 20 to about 80 amino acids. Further provided herein are protein libraries comprising a plurality of proteins, wherein the antibody 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). Further provided herein are protein libraries comprising a plurality of proteins, wherein a length of the VH domain is about 90 to about 100 amino acids. Further provided herein are protein libraries comprising a plurality of proteins, wherein a length of the VL domain is about 90 to about 120 amino acids. Further provided herein are protein libraries comprising a plurality of proteins, wherein the plurality of proteins are used to generate a peptidomimetic library. Further provided herein are protein libraries comprising a plurality of proteins, wherein the protein library comprises antibodies.

[0010] Provided herein are protein libraries comprising a plurality of proteins, wherein each protein of the plurality of proteins comprises a sequence encoding for different PD-1 binding domains, and wherein a length of each PD-1 binding domain is about 20 to about 80 amino acids. Further provided herein are protein libraries comprising a plurality of proteins, wherein the protein library comprises peptides. Further provided herein are protein libraries comprising a plurality of proteins, wherein the protein library comprises immunoglobulins. Further provided herein are protein libraries comprising a plurality of proteins, wherein the protein library comprises

antibodies. Further provided herein are protein libraries comprising a plurality of proteins, wherein the plurality of proteins is used to generate a peptidomimetic library.

[0011] Provided herein are vector libraries comprising a nucleic acid library as described herein. Provided herein are cell libraries comprising a nucleic acid library as described herein. Provided herein are cell libraries comprising a protein library as described herein.

[0012] Provided herein are nucleic acid libraries comprising a plurality of sequences encoding for antibodies or antibody fragments, wherein each sequence of the plurality of sequences comprises a predetermined number of mutations relative to an input sequence; the library comprises at least 30,000 variant sequences; and the at least some of the antibody or antibody fragments bind to PD-1 with a K_D of less than 50 nM. Further provided herein are nucleic acid libraries, wherein the library comprises a CDR sequence of any one of SEQ ID NOs: 1-35. Further provided herein are nucleic acid libraries, wherein the library comprises a CDRH1, CDRH2, or CDRH3 sequence of any one of SEQ ID NOs: 1-35. Further provided herein are nucleic acid libraries, wherein the library comprises at least one sequence encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 10 nM. Further provided herein are nucleic acid libraries, wherein the library comprises at least one sequence encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 5 nM. Further provided herein are nucleic acid libraries, wherein the library comprises at least five sequences encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 10 nM. Further provided herein are nucleic acid libraries, wherein the library comprises at least 50,000 variant sequences. Further provided herein are nucleic acid libraries, wherein the library comprises at least 100,000 variant sequences.

[0013] Provided herein are computerized systems for antibody optimization comprising: (a) a general purpose computer; and (b) a computer readable medium comprising functional modules including instructions for the general purpose computer, wherein said computerized system is configured for operating in a method of: (i) receiving operating instructions, wherein the operating instructions comprise a polynucleotide sequence encoding for an antibody or antibody fragment; (ii) generating an antibody library, wherein the antibody library comprises a plurality of variant sequences of the polynucleotide sequence; and (iii) synthesizing the plurality of variant sequences. Further provided herein are computerized systems for antibody optimization, wherein the antibody library comprises at least 30,000 sequences. Further provided herein are computerized systems for antibody optimization, wherein the antibody library comprises at least 50,000 sequences. Further provided herein are computerized systems for antibody optimization, wherein the antibody library comprises at least 100,000 sequences. Further provided herein are computerized systems for antibody optimization, wherein the system further comprises enriching a subset of the variant

sequences. Further provided herein are computerized systems for antibody optimization, wherein the system further comprises expressing the antibody or antibody fragments corresponding to the variant sequences. Further provided herein are computerized systems for antibody optimization, wherein the polynucleotide sequence is a murine, human, or chimeric antibody sequence. Further provided herein are computerized systems for antibody optimization, wherein the antibody library comprises variant sequences each represented in an amount of no more than 50% of any other variant sequence amount in the antibody library. Further provided herein are computerized systems for antibody optimization, wherein each sequence of the plurality of variant sequences comprises at least one mutation in each CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are computerized systems for antibody optimization, wherein each sequence of the plurality of variant sequences comprises at least two mutations in each CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are computerized systems for antibody optimization, wherein each sequence of the plurality of variant sequences comprises at least one mutation in each CDR of a heavy chain or light chain relative to a germline sequence of the input sequence. Further provided herein are computerized systems for antibody optimization, wherein the antibody library has a theoretical diversity of at least 10^{12} sequences. Further provided herein are computerized systems for antibody optimization, wherein the antibody library has a theoretical diversity of at least 10^{13} sequences.

[0014] Provided herein are methods for optimizing an antibody comprising: (a) providing a polynucleotide sequence encoding for an antibody or antibody fragment; (b) generating an antibody library, wherein the antibody library comprises a plurality of variant sequences of the polynucleotide sequence; and (c) synthesizing the plurality of variant sequences. Further provided herein are methods for optimizing an antibody, wherein the antibody library comprises at least 30,000 sequences. Further provided herein are methods for optimizing an antibody, wherein the antibody library comprises at least 50,000 sequences. Further provided herein are methods for optimizing an antibody, wherein the antibody library comprises at least 100,000 sequences. Further provided herein are methods for optimizing an antibody, wherein the method further comprises enriching a subset of the variant sequences. Further provided herein are methods for optimizing an antibody, wherein the method further comprises expressing the antibody or antibody fragments corresponding to the variant sequences. Further provided herein are methods for optimizing an antibody, wherein the polynucleotide sequence is a murine, human, or chimeric antibody sequence. Further provided herein are methods for optimizing an antibody, wherein the antibody library comprises variant sequences each represented in an amount of no more than 50% of any other variant sequence amount in the antibody library. Further provided herein are methods for

optimizing an antibody, wherein each sequence of the plurality of variant sequences comprises at least one mutation in each CDR of a heavy chain or light chain, relative to the input sequence. Further provided herein are methods for optimizing an antibody, wherein each sequence of the plurality of variant sequences comprises at least two mutations in each CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are methods for optimizing an antibody, wherein each sequence comprises at least one mutation in each CDR of a heavy chain or light chain relative to a germline sequence of the input sequence. Further provided herein are methods for optimizing an antibody, wherein the antibody library has a theoretical diversity of at least 10^{12} sequences. Further provided herein are methods for optimizing an antibody, wherein the antibody library has a theoretical diversity of at least 10^{13} sequences.

[0015] Provided herein are nucleic acid libraries comprising: a plurality of sequences comprising nucleic acids that when translated encode for antibodies or antibody fragments, wherein each of the sequences comprises a predetermined number of mutations within a CDR relative to an input sequence of an antibody; wherein the library comprises at least 50,000 variant sequences, each represented in an amount within 1.5X of a mean frequency; and wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 2.5X higher binding affinity than a binding infinity of the input sequence. Further provided herein are nucleic acid libraries, wherein the library comprises at least 100,000 variant sequences. Further provided herein are nucleic acid libraries, wherein at least some of the sequences encode for an antibody light chain. Further provided herein are nucleic acid libraries, wherein at least some of the sequences encode for an antibody heavy chain. Further provided herein are nucleic acid libraries, wherein each sequence of the plurality of sequences comprises at least one mutation in the CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are nucleic acid libraries, wherein each sequence of the plurality of sequences comprises at least two mutations in the CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are nucleic acid libraries, wherein at least one of the mutations is present in at least two individuals. Further provided herein are nucleic acid libraries, wherein at least one of the mutations is present in at least three individuals. Further provided herein are nucleic acid libraries, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 5X higher binding affinity than a binding infinity of the input sequence. Further provided herein are nucleic acid libraries, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 25X higher binding affinity than a binding infinity of the input sequence. Further provided herein are nucleic acid libraries, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 50X higher binding affinity than a

binding infinity of the input sequence. Further provided herein are nucleic acid libraries, wherein each sequence of the plurality of sequences comprises at least one mutation in the CDR of a heavy chain or light chain relative to a germline sequence of the input sequence. Further provided herein are nucleic acid libraries, wherein the CDR is a CDR1, CDR2, and CDR3 on a heavy chain. Further provided herein are nucleic acid libraries, wherein the CDR is a CDR1, CDR2, and CDR3 on a light chain. Further provided herein are nucleic acid libraries, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having at least 70X higher binding affinity than the input sequence. Further provided herein are nucleic acid libraries, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 50 nM. Further provided herein are nucleic acid libraries, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 25 nM. Further provided herein are nucleic acid libraries, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 10 nM. Further provided herein are nucleic acid libraries, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 5 nM. Further provided herein are nucleic acid libraries, wherein the library comprises a CDR sequence of any one of SEQ ID NOs: 1-6 or 9-70. Further provided herein are nucleic acid libraries, the library comprises a CDRH1, CDRH2, or CDRH3 sequence of any one of SEQ ID NOs: 1-6 or 9-70. Further provided herein are nucleic acid libraries, wherein the library comprises at least one sequence encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 10 nM. Further provided herein are nucleic acid libraries, wherein the library comprises at least one sequence encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 5 nM. Further provided herein are nucleic acid libraries, wherein the library comprises at least five sequences encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 10 nM. Further provided herein are nucleic acid libraries, wherein the library comprises at least 100,000 variant sequences.

[0016] Provided herein are antibodies, wherein the antibody comprises a sequence of any one of SEQ ID NOs: 1-6 or 9-70. Provided herein are antibodies, wherein the antibody comprises a sequence of any one of SEQ ID NOs: 1-6 or 9-34; 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.

[0017] Provided herein are methods of inhibiting PD-1 activity, comprising administering an antibody described herein. Provided herein are methods for treatment of a proliferative disorder, comprising administering to a subject in need thereof an antibody described herein. Further provided herein are methods, wherein the proliferative disorder is cancer. Further provided herein are methods, wherein the cancer is lung, head and neck squamous cell, colorectal, melanoma, liver, classical Hodgkin lymphoma, kidney, gastric, cervical, merkel cell, B-cell lymphoma, or bladder cancer.

[0018] Provided herein are computerized systems for antibody optimization comprising: (a) a general purpose computer; and (b) a computer readable medium comprising functional modules including instructions for the general purpose computer, wherein said computerized system is configured for operating in a method of: (i) receiving operating instructions, wherein the operating instructions comprise a plurality of sequences encoding for an antibody or antibody fragment; (ii) generating a nucleic acid library comprising the plurality of sequences comprising nucleic acids that when translated encode for antibodies or antibody fragments, wherein each of the sequences comprises a predetermined number of mutations within a CDR relative to an input sequence of an antibody; wherein the library comprises at least 50,000 variant sequences, each represented in an amount within 1.5X of a mean frequency; and wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 2.5X higher binding affinity than a binding infinity of the input sequence; and (iii) synthesizing the at least 50,000 variant sequences. Further provided herein are computerized systems for antibody optimization, wherein the nucleic acid library comprises at least 100,000 sequences. Further provided herein are computerized systems for antibody optimization, wherein the system further comprises enriching a subset of the variant sequences. Further provided herein are computerized systems for antibody optimization, wherein the system further comprises expressing the antibody or antibody fragments corresponding to the variant sequences. Further provided herein are computerized systems for antibody optimization, wherein the polynucleotide sequence is a murine, human, or chimeric antibody sequence. Further provided herein are computerized systems for antibody optimization, wherein each sequence of the plurality of variant sequences comprises at least one mutation in a CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are computerized systems for antibody optimization, wherein each sequence of the plurality of variant sequences comprises at least two mutations in a CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are computerized systems for antibody optimization, wherein at

least one sequence when translated encodes for an antibody or antibody fragment having at least 5X higher binding affinity than a binding infinity of the input sequence. Further provided herein are computerized systems for antibody optimization, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 25X higher binding affinity than a binding infinity of the input sequence. Further provided herein are computerized systems for antibody optimization, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 50X higher binding affinity than a binding infinity of the input sequence. Further provided herein are computerized systems for antibody optimization, wherein each sequence of the plurality of variant sequences comprises at least one mutation in a CDR of a heavy chain or light chain relative to a germline sequence of the input sequence. Further provided herein are computerized systems for antibody optimization, wherein the CDR is a CDR1, CDR2, and CDR3 on a heavy chain. Further provided herein are computerized systems for antibody optimization, wherein the CDR is a CDR1, CDR2, and CDR3 on a light chain. Further provided herein are computerized systems for antibody optimization, wherein the antibody library has a theoretical diversity of at least 10^{12} sequences. Further provided herein are computerized systems for antibody optimization, wherein the antibody library has a theoretical diversity of at least 10^{13} sequences.

[0019] Provided herein are methods for optimizing an antibody comprising: (a) providing a plurality of polynucleotide sequences encoding for an antibody or antibody fragment; (b) generating a nucleic acid library comprising the plurality of sequences comprising nucleic acids that when translated encode for antibodies or antibody fragments, wherein each of the sequences comprises a predetermined number of mutations within a CDR relative to an input sequence of an antibody; wherein the library comprises at least 50,000 variant sequences, each represented in an amount within 1.5X of a mean frequency; and wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 2.5X higher binding affinity than a binding infinity of the input sequence; and (c) synthesizing the at least 50,000 variant sequences. Further provided herein are methods for optimizing an antibody, wherein the antibody library comprises at least 100,000 sequences. Further provided herein are methods for optimizing an antibody, wherein the method further comprises enriching a subset of the variant sequences. Further provided herein are methods for optimizing an antibody, wherein the method further comprises expressing the antibody or antibody fragments corresponding to the variant sequences. Further provided herein are methods for optimizing an antibody, wherein the polynucleotide sequence is a murine, human, or chimeric antibody sequence. Further provided herein are methods for optimizing an antibody, wherein each sequence of the plurality of variant sequences comprises

at least one mutation in each CDR of a heavy chain or light chain, relative to the input sequence. Further provided herein are methods for optimizing an antibody, wherein each sequence of the plurality of variant sequences comprises at least two mutations in each CDR of a heavy chain or light chain relative to the input sequence. Further provided herein are methods for optimizing an antibody, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 5X higher binding affinity than a binding infinity of the input sequence. Further provided herein are methods for optimizing an antibody, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 25X higher binding affinity than a binding infinity of the input sequence. Further provided herein are methods for optimizing an antibody, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 50X higher binding affinity than a binding infinity of the input sequence. Further provided herein are methods for optimizing an antibody, wherein each sequence comprises at least one mutation in each CDR of a heavy chain or light chain relative to a germline sequence of the input sequence. Further provided herein are methods for optimizing an antibody, wherein the nucleic acid library has a theoretical diversity of at least 10^{12} sequences. Further provided herein are methods for optimizing an antibody, wherein the nucleic acid library has a theoretical diversity of at least 10^{13} sequences.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] **Figure 1** depicts a workflow for antibody optimization.

[0021] **Figure 2A** depicts a first schematic of an immunoglobulin scaffold.

[0022] **Figure 2B** depicts a second schematic of an immunoglobulin scaffold.

[0023] **Figure 3A** depicts an exemplary sequence of an optimized antibody input sequence in a library, showing the number of mutations in regions of sequence.

[0024] **Figure 3B** shows a workflow for antibody optimization.

[0025] **Figure 4A** depicts read lengths of variable heavy chains after 1-5 rounds of panning.

[0026] **Figure 4B** depicts clonal frequency of variable heavy chains after 1-5 rounds of panning.

[0027] **Figure 4C** depicts clonal accumulation of variable heavy chains after 1-5 rounds of panning.

[0028] **Figure 4D** is a graph of the number of mutations vs. different panning conditions.

[0029] **Figure 4E** is a graph of sequence analysis of enriched clones for binding to PD-1.

[0030] **Figure 5A** is a plot anti-scFv ELISA vs. enrichment at round five of panning.

[0031] **Figure 5B** is a graph of anti-scFv ELISA binding to PD-1.

[0032] **Figure 6A** depicts sequences of optimized IgGs which bind to comparator PD-1.

[0033] **Figure 6B** depicts an increase in affinity of an optimized antibody (4.5 nM) as compared to the parental antibody (330 nM).

[0034] **Figures 6C-6D** depict sequence alignment of CDR's.

[0035] **Figure 6E** depicts dose-dependent phage PD-1 ELISA.

[0036] **Figure 6F** depicts an iso-affinity plot.

[0037] **Figure 7A** is a graph of PD-1/PDL-1 blockade assay of optimized IgGs.

[0038] **Figure 7B** shows binding affinity and potency of optimized IgGs.

[0039] **Figure 7C** is a graph of PD-1/PDL-1 blockade IC50 (nM, y-axis) versus SPR monovalent binding affinity (Kd (nM), x-axis) of optimized IgGs.

[0040] **Figure 7D** is a graph of BVP scores (y-axis) of several IgGs (x-axis).

[0041] **Figure 8** presents a diagram of steps demonstrating an exemplary process workflow for gene synthesis as disclosed herein.

[0042] **Figure 9** illustrates an example of a computer system.

[0043] **Figure 10** is a block diagram illustrating an architecture of a computer system.

[0044] **Figure 11** 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).

[0045] **Figure 12** is a block diagram of a multiprocessor computer system using a shared virtual address memory space.

DETAILED DESCRIPTION

[0046] 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.

[0047] **Definitions**

[0048] 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.

[0049] 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.

[0050] 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.

[0051] 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.

[0052] Antibody Optimization

[0053] Provided herein are methods, compositions, and systems for the optimization of antibodies. Antibodies are in some instances optimized by the design of in-silico libraries comprising variant sequences of an input antibody sequence (**FIG. 1**). Input sequences **100** are in some instances modified in-silico **102** with one or more mutations to generate libraries of optimized sequences **103**. In some instances, such libraries are synthesized, cloned into expression vectors, and translation products (antibodies) evaluated for activity. In some instances, fragments of sequences are synthesized and subsequently assembled. In some instances, expression vectors are used to display and enrich desired antibodies, such as phage display. Selection pressures used during enrichment in some instances includes binding affinity, toxicity, immunological tolerance, stability, or other factor. Such expression vectors allow antibodies with specific properties to be selected ("panning"), and subsequent propagation or amplification of such sequences enriches the library with these sequences. Panning rounds can be repeated any number of times, such as 1, 2, 3, 4, 5, 6, 7, or more than 7 rounds. Sequencing at one or more rounds is in some instances used to identify which sequences **105** have been enriched in the library.

[0054] Described herein are methods and systems of in-silico library design. For example, an antibody or antibody fragment sequence is used as input. Any antibody sequence is in some instances used for input in to the methods and systems described herein. A database **102** comprising known mutations from an organism is queried **101**, and a library **103** of sequences comprising combinations of these mutations are generated. In some instances, antibodies described herein comprise CDR regions. In some instances, known mutations from CDRs are used to build the sequence library. Filters **104**, or exclusion criteria, are in some instances used to select specific types of variants for members of the sequence library. For example, sequences having a mutation

are added if a minimum number of organisms in the database have the mutation. In some instances, additional CDRs are specified for inclusion in the database. In some instances, specific mutations or combinations of mutations are excluded from the library (e.g., known immunogenic sites, structure sites, etc.). In some instances, specific sites in the input sequence are systematically replaced with histidine, aspartic acid, glutamic acid, or combinations thereof. In some instances, the maximum or minimum number of mutations allowed for each region of an antibody are specified. Mutations in some instances are described relative to the input sequence or the input sequence's corresponding germline sequence. For example, sequences generated by the optimization comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or more than 16 mutations from the input sequence. In some instances, sequences generated by the optimization comprise no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or no more than 18 mutations from the input sequence. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or about 18 mutations relative to the input sequence. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a first CDR region. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a second CDR region. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a third CDR region. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a first CDR region of a heavy chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a second CDR region of a heavy chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a third CDR region of a heavy chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a first CDR region of a light chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a second CDR region of a light chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the input sequence in a third CDR region of a light chain. In some instances, a first CDR region is CDR1. In some instances, a second CDR region is CDR2. In some instances, a third CDR region is CDR3. In-silico antibodies libraries are in some instances synthesized, assembled, and enriched for desired sequences.

[0055] The germline sequences corresponding to an input sequence may also be modified to generate sequences in a library. For example, sequences generated by the optimization methods described herein comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or more than 16

mutations from the germline sequence. In some instances, sequences generated by the optimization comprise no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or no more than 18 mutations from the germline sequence. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or about 18 mutations relative to the germline sequence.

[0056] Provided herein are methods, systems, and compositions for antibody optimization, wherein the input sequence comprises mutations in an antibody region. Exemplary regions of the antibody 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). In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a first CDR region. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a second CDR region. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a third CDR region. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a first CDR region of a heavy chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a second CDR region of a heavy chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a third CDR region of a heavy chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a first CDR region of a light chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a second CDR region of a light chain. In some instances, sequences generated by the optimization comprise about 1, 2, 3, 4, 5, 6, or 7 mutations from the germline sequence in a third CDR region of a light chain. In some instances, a first CDR region is CDR1. In some instances, a second CDR region is CDR2. In some instances, a third CDR region is CDR3.

[0057] *Antibody Libraries*

[0058] Provided herein are libraries generated from antibody optimization methods described herein. Antibodies described herein result in improved functional activity, structural stability, expression, specificity, or a combination thereof.

[0059] 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 an antibody, wherein the antibody 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 an antibody, wherein the antibody 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.

[0060] In some embodiments, libraries comprise immunoglobulins that are adapted to the species of an intended therapeutic target. Generally, these methods include "mammalization" and comprises methods for transferring donor antigen-binding information to a less immunogenic mammal antibody acceptor to generate useful therapeutic treatments. In some instances, the mammal is mouse, rat, equine, sheep, cow, primate (*e.g.,* chimpanzee, baboon, gorilla, orangutan, monkey), dog, cat, pig, donkey, rabbit, and human. In some instances, provided herein are libraries and methods for felinization and caninization of antibodies.

[0061] "Humanized" forms of non-human antibodies can be chimeric antibodies that contain minimal sequence derived from the non-human antibody. A humanized antibody is generally a human antibody (recipient antibody) in which residues from one or more CDRs are replaced by residues from one or more CDRs of a non-human antibody (donor antibody). The donor antibody can be any suitable non-human antibody, such as a mouse, rat, rabbit, chicken, or non-human primate antibody having a desired specificity, affinity, or biological effect. In some instances, selected framework region residues of the recipient antibody are replaced by the corresponding framework region residues from the donor antibody. Humanized antibodies may also comprise residues that are not found in either the recipient antibody or the donor antibody. In some instances, these modifications are made to further refine antibody performance.

[0062] "Caninization" can comprise a method for transferring non-canine antigen-binding information from a donor antibody to a less immunogenic canine antibody acceptor to generate treatments useful as therapeutics in dogs. In some instances, caninized forms of non-canine antibodies provided herein are chimeric antibodies that contain minimal sequence derived from non-canine antibodies. In some instances, caninized antibodies are canine antibody sequences ("acceptor" or "recipient" antibody) in which hypervariable region residues of the recipient are replaced by hypervariable region residues from a non-canine species ("donor" antibody) such as mouse, rat, rabbit, cat, dogs, goat, chicken, bovine, horse, llama, camel, dromedaries, sharks, non-

human primates, human, humanized, recombinant sequence, or an engineered sequence having the desired properties. In some instances, framework region (FR) residues of the canine antibody are replaced by corresponding non-canine FR residues. In some instances, caninized antibodies include residues that are not found in the recipient antibody or in the donor antibody. In some instances, these modifications are made to further refine antibody performance. The caninized antibody may also comprise at least a portion of an immunoglobulin constant region (Fc) of a canine antibody.

[0063] "Felinization" can comprise a method for transferring non-feline antigen-binding information from a donor antibody to a less immunogenic feline antibody acceptor to generate treatments useful as therapeutics in cats. In some instances, felinized forms of non-feline antibodies provided herein are chimeric antibodies that contain minimal sequence derived from non-feline antibodies. In some instances, felinized antibodies are feline antibody sequences ("acceptor" or "recipient" antibody) in which hypervariable region residues of the recipient are replaced by hypervariable region residues from a non-feline species ("donor" antibody) such as mouse, rat, rabbit, cat, dogs, goat, chicken, bovine, horse, llama, camel, dromedaries, sharks, non-human primates, human, humanized, recombinant sequence, or an engineered sequence having the desired properties. In some instances, framework region (FR) residues of the feline antibody are replaced by corresponding non-feline FR residues. In some instances, felinized antibodies include residues that are not found in the recipient antibody or in the donor antibody. In some instances, these modifications are made to further refine antibody performance. The felinized antibody may also comprise at least a portion of an immunoglobulin constant region (Fc) of a felinize antibody.

[0064] Methods as described herein may be used for optimization of libraries encoding a non-immunoglobulin. In some instances, the libraries comprise antibody mimetics. Exemplary antibody mimetics include, but are not limited to, anticalins, affilins, affibody molecules, affimers, affitins, alphabodies, avimers, atrimers, DARPins, fynomers, Kunitz domain-based proteins, monobodies, anticalins, knottins, armadillo repeat protein-based proteins, and bicyclic peptides.

[0065] Libraries described herein comprising nucleic acids encoding for an antibody comprise variations in at least one region of the antibody. 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).

[0066] Methods described herein provide for synthesis of libraries comprising nucleic acids encoding an antibody, 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 antibody 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.

[0067] In some instances, the at least one region of the antibody 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).

[0068] Provided herein are libraries comprising nucleic acids encoding for antibodies, 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 antibody 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.

[0069] Libraries comprising nucleic acids encoding for antibodies 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 antibodies 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.

[0070] A number of variant sequences for the at least one region of the antibody 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). *See* **FIG. 2A**. 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.

[0071] Variant sequences for the at least one region of the antibody, 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.

[0072] Following synthesis of antibody libraries, antibody libraries may be used for screening and analysis. For example, antibody 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, antibody 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. In some instances, antibody libraries are displayed on the surface of a cell or phage. In some instances, antibody libraries are enriched for sequences with a desired activity using phage display.

[0073] In some instances, the antibody 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 antibody libraries are assayed for antibody 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.

[0074] Antibodies or IgGs generated by methods as described herein comprise improved binding affinity. In some instances, the antibody comprises a binding affinity (e.g., kD) of less than 1 nM, less than 1.2 nM, less than 2 nM, less than 5 nM, less than 10 nM, less than 11 nm, less than 13.5 nM, less than 15 nM, less than 20 nM, less than 25 nM, or less than 30 nM. In some instances, the antibody comprises a kD of less than 1 nM. In some instances, the antibody comprises a kD of less than 1.2 nM. In some instances, the antibody comprises a kD of less than 2 nM. In some instances, the antibody comprises a kD of less than 5 nM. In some instances, the antibody comprises a kD of less than 10 nM. In some instances, the antibody comprises a kD of less than 13.5 nM. In some instances, the antibody comprises a kD of less than 15 nM. In some instances, the antibody comprises a kD of less than 20 nM. In some instances, the antibody comprises a kD of less than 25 nM. In some instances, the antibody comprises a kD of less than 30 nM.

[0075] In some instances, the affinity of antibodies or IgGs generated by methods as described herein is at least or about 1.5x, 2.0x, 5x, 10x, 20x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x,

or more than 200x improved binding affinity as compared to a comparator antibody. In some instances, the affinity of antibodies or IgGs generated by methods as described herein is at least or about 1.5x, 2.0x, 5x, 10x, 20x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, or more than 200x improved function as compared to a comparator antibody. In some instances, the comparator antibody is an antibody with similar structure, sequence, or antigen target.

[0076] *Expression Systems*

[0077] Provided herein are libraries comprising nucleic acids encoding for antibody comprising 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.

[0078] Provided herein are libraries comprising nucleic acids encoding for antibody comprising 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.

[0079] 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.

[0080] 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.

[0081] Described herein are nucleic acid libraries that are expressed in a vector to generate a construct comprising an antibody. 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.

[0082] Provided herein are libraries comprising nucleic acids encoding for antibodies, 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 glucoronidase (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.

[0083] *PD-1 Libraries*

[0084] Provided herein are methods and compositions relating to program cell death protein 1 (PD-1) binding libraries comprising nucleic acids encoding for a PD-1 antibody. Such methods and compositions in some instances are generated by the antibody optimization methods and systems described herein. Antibodies as described herein can stably support a PD-1 binding domain. The PD-1 binding domain may be designed based on surface interactions of a PD-1 ligand and PD-1. 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 PD-1 signaling. In some instances, an antibody described herein comprises a CDRH1 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 80% identical to a CDRH1 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 85% identical to a CDRH1 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 90% identical to a CDRH1 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 95% identical to a CDRH1 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a CDRH2 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 80% identical to a CDRH2 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 85% identical to a CDRH2 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 90% identical to a CDRH2 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 95% identical to a CDRH2 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a CDRH3 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 80% identical to a CDRH3 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 85% identical to a CDRH3 sequence

of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 90% identical to a CDRH3 sequence of any one of SEQ ID NOs: 1-35. In some instances, an antibody described herein comprises a sequence that is at least 95% identical to a CDRH3 sequence of any one of SEQ ID NOs: 1-35.

[0085] The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

[0086] The term "homology" or "similarity" between two proteins is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one protein sequence to the second protein sequence. Similarity may be determined by procedures which are well-known in the art, for example, a BLAST program (Basic Local Alignment Search Tool at the National Center for Biological Information).

[0087] Provided herein are libraries comprising nucleic acids encoding for PD-1 antibodies. Antibodies described herein allow for improved stability for a range of PD-1 binding domain encoding sequences. In some instances, the PD-1 binding domain encoding sequences are determined by interactions between the PD-1 ligand and PD-1.

[0088] Sequences of PD-1 binding domains based on surface interactions between a PD-1 ligand and PD-1 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/).

[0089] Described herein are PD-1 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. In some instances, homologous sequences are identified in the same organism, across individuals.

[0090] Following identification of PD-1 binding domains, libraries comprising nucleic acids encoding for the PD-1 binding domains may be generated. In some instances, libraries of PD-1 binding domains comprise sequences of PD-1 binding domains designed based on conformational ligand interactions, peptide ligand interactions, small molecule ligand interactions, extracellular domains of PD-1s, or antibodies that target PD-1s. Libraries of PD-1 binding domains may be translated to generate protein libraries. In some instances, libraries of PD-1 binding domains are translated to generate peptide libraries, immunoglobulin libraries, derivatives thereof, or combinations thereof. In some instances, libraries of PD-1 binding domains are translated to generate protein libraries that are further modified to generate peptidomimetic libraries. In some instances, libraries of PD-1 binding domains are translated to generate protein libraries that are used to generate small molecules.

[0091] Methods described herein provide for synthesis of libraries of PD-1 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 PD-1 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 PD-1 binding domain. In some instances, the variant library comprises sequences encoding for variation of multiple codons in a PD-1 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.

[0092] Methods described herein provide for synthesis of libraries comprising nucleic acids encoding for the PD-1 binding domains, wherein the libraries comprise sequences encoding for variation of length of the PD-1 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.

[0093] Following identification of PD-1 binding domains, antibodies may be designed and synthesized to comprise the PD-1 binding domains. Antibodies comprising PD-1 binding domains may be designed based on binding, specificity, stability, expression, folding, or downstream activity. In some instances, the antibodies comprising PD-1 binding domains enable contact with the PD-1. In some instances, the antibodies comprising PD-1 binding domains enables high affinity binding with the PD-1. Exemplary amino acid sequences of PD-1 binding domains comprise any one of SEQ ID NOs: 1-70.

[0094] In some instances, the PD-1 antibody comprises a binding affinity (e.g., kD) to PD-1of less than 1 nM, less than 1.2 nM, less than 2 nM, less than 5 nM, less than 10 nM, less than 11 nm, less than 13.5 nM, less than 15 nM, less than 20 nM, less than 25 nM, or less than 30 nM. In some instances, the PD-1 antibody comprises a kD of less than 1 nM. In some instances, the PD-1 antibody comprises a kD of less than 1.2 nM. In some instances, the PD-1 antibody comprises a kD of less than 2 nM. In some instances, the PD-1 antibody comprises a kD of less than 5 nM. In some instances, the PD-1 antibody comprises a kD of less than 10 nM. In some instances, the PD-1 antibody comprises a kD of less than 13.5 nM. In some instances, the PD-1 antibody comprises a kD of less than 15 nM. In some instances, the PD-1 antibody comprises a kD of less than 20 nM. In some instances, the PD-1 antibody comprises a kD of less than 25 nM. In some instances, the PD-1 antibody comprises a kD of less than 30 nM.

[0095] In some instances, the affinity the PD-1 antibody generated by methods as described herein is at least or about 1.5x, 2.0x, 5x, 10x, 20x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, or more than 200x improved binding affinity as compared to a comparator antibody. In some instances, the PD-1 antibody generated by methods as described herein is at least or about 1.5x, 2.0x, 5x, 10x, 20x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, or more than 200x improved function as compared to a comparator antibody. In some instances, the comparator antibody is an antibody with similar structure, sequence, or antigen target.

[0096] Provided herein are PD-1 binding libraries comprising nucleic acids encoding for antibodies comprising PD-1 binding domains comprise variation in domain type, domain length, or residue variation. In some instances, the domain is a region in the antibody comprising the PD-1 binding domains. For example, the region is the VH, CDR-H3, or VL domain. In some instances, the domain is the PD-1 binding domain.

[0097] Methods described herein provide for synthesis of a PD-1 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 PD-1 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 PD-1 binding domain. For example, at least one single codon of a PD-1 binding domain 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 PD-1 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.

[0098] Methods described herein provide for synthesis of a PD-1 binding library of nucleic acids each encoding for a predetermined variant of at least one predetermined reference nucleic acid sequence, wherein the PD-1 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 PD-1 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.

[0099] Provided herein are PD-1 binding libraries comprising nucleic acids encoding for antibodies comprising PD-1 binding domains, wherein the PD-1 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 PD-1 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.

[00100] PD-1 binding libraries comprising nucleic acids encoding for antibodies comprising PD-1 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.

[00101] PD-1 binding libraries comprising de novo synthesized variant sequences encoding for antibodies comprising PD-1 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 are de novo synthesized for a PD-1 binding domain. 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.

[00102] PD-1 binding libraries comprising de novo synthesized variant sequences encoding for antibodies comprising PD-1 binding domains comprise improved diversity. In some instances, variants include affinity maturation variants. Alternatively or in combination, variants include variants in other regions of the antibody 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 PD-1 binding libraries is least or about 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{13}, 10^{14} or more than 10^{14} non-identical sequences.

[00103] Following synthesis of PD-1 binding libraries comprising nucleic acids encoding antibodies comprising PD-1 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, PD-1 binding libraries comprise nucleic acids encoding antibodies comprising PD-1 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.

[00104] Diseases and Disorders

[00105] Provided herein are PD-1 binding libraries comprising nucleic acids encoding for antibodies comprising PD-1 binding domains may have therapeutic effects. In some instances, the PD-1 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, the cancer is lung, head and neck squamous cell, colorectal, melanoma, liver, classical Hodgkin lymphoma, kidney, gastric, cervical, merkel cell, B-cell lymphoma, or bladder cancer. In some instances, the cancer is a MSI-H/dMMR cancer. In some instances, an inhibitor of PD-1 programmed cell death protein 1 as described herein is used for treatment of a metabolic disorder. 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: 1-70. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat lung cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat head and neck squamous cell cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat colorectal cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat melanoma. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat liver cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat classical Hodgkin lymphoma. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat kidney cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used

to treat gastric cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat cervical cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat merkel cell cancer. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat B-cell lymphoma. In some instances, a sequence of any one of SEQ ID NOs: 1-70 is used to treat bladder cancer.

[00106] Variant Libraries

[00107] *Codon variation*

[00108] 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 1** provides a listing of each codon possible (and the representative amino acid) for a variant site.

Table 1. 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				

[00109] 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.

[00110] Highly Parallel Nucleic Acid Synthesis

[00111] 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.

[00112] 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.

[00113] 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.

[00114] 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.

[00115] 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.

[00116] 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.

[00117] 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.

[00118] **Substrates**

[00119] 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.

[00120] 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.

[00121] 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.

[00122] 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 um. 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 um. 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 um. 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 um. In some cases, each locus has a width of about 0.5-100, 0.5-50, 10-75, or 0.5-50 um.

[00123] 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 um. 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 um. 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.

[00124] 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.

[00125] *Surface materials*

[00126] 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; fuse 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.

[00127] Surface Architecture

[00128] 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.

[00129] 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.

[00130] 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 mm^2.

[00131] 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.

[00132] 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 um. In some cases, the height of a well is less than about 1000, 900, 800, 700, or 600 um.

[00133] 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 um. In some cases, the height of a channel is less than 100, 80, 60, 40, or 20 um.

[00134] 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 um, for example, about 90, 80, 70, 60, 50, 40, 30, 20 or 10 um. 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 um. 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 um.

[00135] Surface Modifications

[00136] 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.

[00137] 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.

[00138] 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.

[00139] 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 alkoxysilane molecule. Methods and functionalizing agents are described in U.S. Patent No. 5474796, which is herein incorporated by reference in its entirety.

[00140] 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 alkoxysilane 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 alkoxysilanes are classified according to their organic functions.

[00141] **Polynucleotide Synthesis**

[00142] 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.

[00143] 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).

[00144] 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 I_2 /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 I_2/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.

[00145] 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).

[00146] 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.

[00147] 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.

[00148] 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.

[00149] 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.

[00150] 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.

[00151] 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.

[00152] 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.

[00153] 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.

[00154] Referring to the Figures, **FIG. 8** 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.

[00155] 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.

[00156] 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 **802**. In some instances, polynucleotides are cleaved from the surface at this stage. Cleavage includes gas cleavage, *e.g.*, with ammonia or methylamine.

[00157] 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 **803**. Prior to or after the sealing **804** 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 **805**. Once released, fragments of single stranded polynucleotides hybridize in order to span an entire long range sequence of DNA. Partial hybridization **805** is possible because each synthesized polynucleotide is designed to have a small portion overlapping with at least one other polynucleotide in the pool.

[00158] 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 **806**.

[00159] After PCA is complete, the nanoreactor is separated from the device **807** and positioned for interaction with a device having primers for PCR **808**. After sealing, the nanoreactor is subject to PCR **809** and the larger nucleic acids are amplified. After PCR **810**, the nanochamber is opened **811**, error correction reagents are added **812**, the chamber is sealed **813** and an error correction reaction occurs to remove mismatched base pairs and/or strands with poor complementarity from the double stranded PCR amplification products **814**. The nanoreactor is opened and separated **815**. Error corrected product is next subject to additional processing steps, such as PCR and molecular bar coding, and then packaged **822** for shipment **823**.

[00160] 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 **816**, sealing the wafer to a chamber containing error corrected amplification product **817**, and performing an additional round of amplification **818**. The nanoreactor is opened **819** and the products are pooled **820** and sequenced **821**. After an acceptable quality control determination is made, the packaged product **822** is approved for shipment **823**.

[00161] In some instances, a nucleic acid generate by a workflow such as that in **FIG. 8** 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 **802**.

[00162] **Computer systems**

[00163] 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.

[00164] The computer system **900** illustrated in **FIG. 9** may be understood as a logical apparatus that can read instructions from media **911** and/or a network port **905**, which can optionally be connected to server **909** having fixed media **912**. The system, such as shown in **FIG. 9** can include a CPU **901**, disk drives **903**, optional input devices such as keyboard **915** and/or mouse **916** and optional monitor **907**. 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 **922** as illustrated in **FIG. 9**.

[00165] **FIG. 10** is a block diagram illustrating a first example architecture of a computer system **1000** that can be used in connection with example instances of the present disclosure. As depicted in **FIG. 10**, the example computer system can include a processor **1002** for processing instructions. Non-limiting examples of processors include: Intel XeonTM processor, AMD OpteronTM processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0TM processor, ARM Cortex-A8 Samsung S5PC100TM processor, ARM Cortex-A8 Apple A4TM processor, Marvell PXA 930TM 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.

[00166] As illustrated in **FIG. 10**, a high speed cache **1004** can be connected to, or incorporated in, the processor **1002** to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor **1002**. The processor **1002** is connected to a north bridge **1006** by a processor bus **1008**. The north bridge **1006** is connected to random access memory (RAM) **1010** by a memory bus **1012** and manages access to the RAM **1010** by the processor **1002**. The north bridge **1006** is also connected to a south bridge **1014** by a chipset bus **1016**. The south bridge **1014** is, in turn, connected to a peripheral bus **1018**. 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 **1018**. 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 **1000** can include an accelerator card **1022** attached to the peripheral bus **1018**. 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.

[00167] Software and data are stored in external storage **1024** and can be loaded into RAM **1010** and/or cache **1004** for use by the processor. The system **1000** 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 **1000** also includes network interface cards (NICs) **1020** and **1021** 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.

[00168] **FIG. 11** is a diagram showing a network **1100** with a plurality of computer systems **1102a**, and **1102b**, a plurality of cell phones and personal data assistants **1102c**, and Network Attached Storage (NAS) **1104a**, and **1104b**. In example instances, systems **1102a**, **1102b**, and **1102c** can manage data storage and optimize data access for data stored in Network Attached Storage (NAS) **1104a** and **1104b**. A mathematical model can be used for the data and be evaluated using distributed parallel processing across computer systems **1102a**, and **1102b**, and cell phone and personal data assistant systems **1102c**. Computer systems **1102a**, and **1102b**, and cell phone and personal data assistant systems **1102c** can also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) **1104a** and **1104b**. **FIG. 11** 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.

[00169] **FIG. 12** is a block diagram of a multiprocessor computer system using a shared virtual address memory space in accordance with an example instance. The system includes a plurality of processors **1202a-f** that can access a shared memory subsystem **1204**. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) **1206a-f** in the memory subsystem **1204**. Each MAP **1206a-f** can comprise a memory **1208a-f** and one or more field programmable gate arrays (FPGAs) **1210a-f**. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs **1210a-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 **1208a-f**, allowing it to execute tasks independently of, and asynchronously from the respective microprocessor **1202a-f**. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[00170] 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.

[00171] 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. 10**, 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 **1022** illustrated in **FIG. 10**.

[00172] 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

[00173] 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.

[00174] **Example 1: Functionalization of a device surface**

[00175] 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% H_2SO_4 and 10% H_2O_2 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 N_2. The device was subsequently soaked in NH_4OH (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 O_2. A SAMCO PC-300 instrument was used to plasma etch O_2 at 250 watts for 1 min in downstream mode.

[00176] 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 O_2 plasma etch at 250 watts for 1 min.

[00177] 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 N_2. The functionalized surface was activated to serve as a support for polynucleotide synthesis.

[00178] **Example 2: Synthesis of a 50-mer sequence on an oligonucleotide synthesis device**

[00179] 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-TRIETHOXYSILYLPROPYL)-4-HYDROXYBUTYRAMIDE (Gelest) was used to synthesize an exemplary polynucleotide of 50 bp ("50-mer polynucleotide") using polynucleotide synthesis methods described herein.

[00180] The sequence of the 50-mer was as described. 5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTTTTTT3', 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.

[00181] The synthesis was done using standard DNA synthesis chemistry (coupling, capping, oxidation, and deblocking) according to the protocol in **Table 2** and an ABI synthesizer.

Table 2: Synthesis protocols

	Table 2	
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	Acetonitrile System Flush	4

Table 2		
General DNA Synthesis Process Name	**Process Step**	**Time (sec)**
Flow)	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
OXIDATION (Oxidizer Flow)	Oxidizer to Flowcell	18
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

[00182] 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.

[00183] 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-methylimidizole 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.

[00184] Example 3: Synthesis of a 100-mer sequence on an oligonucleotide synthesis device

[00185] 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' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATG CTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTTT3', where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes) on two different silicon chips, the first one uniformly functionalized with N-(3-TRIETHOXYSILYLPROPYL)-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.

[00186] All ten samples from the two chips were further PCR amplified using a forward (5'ATGCGGGGTTCTCATCATC3') and a reverse (5'CGGGATCCTTATCGTCATCG3') 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 ºC, 30 sec

98 ºC, 10 sec; 63 ºC, 10 sec; 72 ºC, 10 sec; repeat 12 cycles

72 ºC, 2min

[00187] 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 3** 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 3: 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%
8	1/1769 bp	99.94%
9	1/854 bp	99.88%

Spot	Error rate	Cycle efficiency
10	1/1451 bp	99.93%

[00188] 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.

[00189] **Table 4** summarizes error characteristics for the sequences obtained from the polynucleotides samples from spots 1-10.

Table 4: Error characteristics

Sample ID/Spot no.	OSA_0046/1	OSA_0047/2	OSA_0048/3	OSA_0049/4	OSA_0050/5	OSA_0051/6	OSA_0052/7	OSA_0053/8	OSA_0054/9	OSA_0055/10
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 Minus Primer	MP Err: ~1 in 763	MP Err: ~1 in 824	MP Err: ~1 in 780	MP Err: ~1 in 429	MP Err: ~1 in 1525	MP Err: ~1 in 1615	MP Err: ~1 in 531	MP Err: ~1 in 1769	MP Err: ~1 in 854	MP Err: ~1 in 1451

Error Rate										

[00190] **Example 4: Antibody optimization**

[00191] *Library Generated from Parent Sequence*

[00192] Antibody sequences targeting PD-1 were designed by generating, in-silico, a library comprising mutations from twelve individuals. Heavy chain and light chain mutational space was derived from the parent sequence and the closest germline sequence to generate an NGS database. The NGS database comprised sequences of light chain CDR1-3 and heavy chain CDR1-3 comprising mutations as compared to the parent reference sequence or germline sequence. All CDR sequences were represented in two or more individuals from the NGS database. The input sequence is shown in **FIG. 3A**. The library contained $5.9x10^7$ different heavy chains, and $2.9x10^6$ light chains (**FIG. 3B**).

[00193] *Bead-based Selections*

[00194] C-terminal biotinylated PD-1 antigen was bound to streptavidin-coated magnetic beads for five rounds of selections. Bead-binding variants were depleted between each round. Stringency of selections was increased with each subsequent round, and enrichment ratios track on-target binding.

[00195] *ELISA and Next Generation Sequencing*

[00196] Constructs expressing combinations of heavy and light chains were synthesized and subjected to phage display to identify improved binders to PD-1. The pool was sequenced with 10 million reads, and 400,000 unique clones were identified. Read length distributions were highly uniform (**FIG. 4A**). Clonal frequency and accumulation were measured at each panning round (**FIGS. 4B** and **4C**). Four different stringency conditions were used for panning (**FIG. 4D**). High-stringency selections enriched for the same pool of binders across repeat selections. Low-stringency selections recovered 44 of 70 (63%) of the same clones as from high-stringency selections, with a broader range of low-affinity binders. The majority of scFv binders were enriched in round 5 (**FIG. 5A**). Greater than 90% (68/75) of clones were present at 5x over background in an ELISA experiment measuring scFv binding to PD-1, and were enriched to greater than 0.01%.

[00197] Five rounds of selections were completed with three different initial selection conditions. Clone enrichment was tracked through each successive round by NGS. Sequences that were enriched for off-target or background binders are removed. Approximately 1000 clones represented in Round 5 were enriched to greater than 0.01% of population. Sequence analysis show that the vast majority of enriched clones (>95%) for binding to PD-1 were equally captured across the different selection conditions (**FIG. 4E**).

[00198] *High-throughput IgG characterization*

[00199] Clones were transiently transfected in Expi293, and purified by Kingfisher and Hamilton automation decks. Yield and purity were confirmed by Perkin Elmer Labchip and analytical HPLC. The Carterra LSA system was used to assess binding affinity and epitope binning of >170 IgG variants (data not shown).

[00200] Optimized IgGs bound with similar or improved affinity to comparator antibodies having a sequence corresponding to pembrolizumab (Comparator 1) and a sequence corresponding to nivolumab (Comparator 2) (data not shown). scFV binding to PD-1 was also measured as seen in **FIG. 5B**. Sequences of optimized antibodies contained fewer germline mutations than comparator antibodies having a sequence corresponding to pembrolizumab (Comparator 1) and a sequence corresponding to nivolumab (Comparator 2) (**FIG. 6A** and **Table 5A**). Light chains of the optimized antibodies were highly diverse, with greater than 90% of clones containing unique light chains that were never repeated (**Table 5B**). Optimized IgGs demonstrated 100x improvement in monovalent binding affinity compared to the parental sequence. The PD1-1 clone bound to PD-1 with a K_D of 4.52 nM, while several others show binding affinities < 10 nM. **FIG. 6B** shows the affinity increase following optimization using methods described herein. These high-affinity binders each comprised unique CDRH3 and were not clustered by sequence lineage. Diversity across the various CDRs was also observed as seen in **FIGS. 6C-6D**. PD-1 antibodies exhibited improved binding affinity compared to wild-type (**FIGS. 6E-6F** and **Table 5C**).

Table 5A. PD-1 Variant Sequences

SEQ ID NO.	Clone	Sequence
1	PD1-1	YTFSHYGISGWVSAYSGNTRYACARDATYSSGSGYW
2	PD1-15	YTFTDYGVSYTISPYSGKTDYACARDADYSIGSGYW
3	PD1-91	YTFADYGISGWISAYNGNTNYACARDWDYSSGSGYW
4	PD1-2	YSFSNYGISGWVSAYSGNTNYACARDADYASGSGYW
5	PD1-7	YTFTHYGITGWISAYSGSTNYACARDADYSIGSGYW
6	PD1-75	YPFTNYGISGWISPYSGNTNYACARDFDYSSGSGYW
7	Sequence corresponding to nivolumab	YTFTNYYMYGGINPSNGGTNFNCARRDYRFDMGFDY
8	Sequence corresponding to pembrolizumab	ITFSNSGMHVIWWYDGSKRYYACATNDDY
9	PD1-60	YDFSSYGISGWISAYSGNTRYACARDADYSVGSGYW
10	PD1-88	YTFNRYGISGWISAYSGNTRYACARDADYASGSGYW
11	PD1-4	YTFSTSGISGWISAYNGNTNYACARDFDYSSGSGYW
12	PD1-86	YTFDDYGISGWISAYNGNTNYACARDATYSSGSGYW
13	PD1-54	YAFSNYGISGWISPYSGNTKYACARDADYSIGSGYW

14	PD1-14	YPFTNYGISGWTSAYSGNTNYACARDFDYSSGSGYW
15	PD1-82	YRFTSYGISGWISAYNGNTNYACARDWDYSSGSGYW
16	PD1-10	YRFTSYGISGWISAYNGNTNYACARDADYASGSGYW
17	PD1-11	YTFSNYGFSGWISPYSGNTKYACARDADYSIGSGYW
18	PD1-76	YRFTSYGISGWISAYNGNTNYACARDFDYSSGSGYW
19	PD1-12	YTFNTYGVSGWISAYSGNTRYACARDFDYSSGSGYW
20	PD1-42	YTFNNFGISGWISAYNGNTDYACARDMDYSSGSGYW
21	PD1-43	YTFSNYGISGWISAYSGNTKYACARDADYASGSGYW
22	PD1-17	YTFSKYGISGWISAYSGHTDYACARDFDYSSGSGYW
23	PD1-89	YSFTKYGISGWISAYSGNTRYACARDADYSIGSGYW
24	PD1-87	YTFDRYGISGWISAYSGNTRYACARDVDYSSGSGYW
25	PD1-48	YTFSHYGISGWISPYSGNTKYACARDADYWSGSGYW
26	PD1-59	YDFSSYGISGWISAYSGYTNYACARDAVYSSGSGYW
27	PD1-18	YTFNNYGISGWISAYSGHTDYACARDWDYSSGSGYW
28	PD1-30	YTFNNYGVSGWVSAYSGNTNYACARDADYASGSGYW
29	PD1-23	YPFTSYGISGWISAYNGNTNYACARDVDYSSGSGYW
30	PD1-25	YSFTRYGISGWISPYSGNTKYACARDADYGSGSGYW
31	PD1-8	YTFSHYGISGWISPYSGNTKYACARDNDYSSGSGYW
32	PD1-36	YTFSSYGVSGWISAYSGSTNYACARDAVYSSGSGYW
33	PD1-50	YTFRNYGISGWTSAYNGNTNYACARDATYSSGSGYW
34	PD1-27	YPFTNYGITGWISAYNGNTNYACARDWDYSSGSGYW
35	Parental	YRFTSYGISGWISAYNGNTNYACARDADYSSGSGYW

Table 5B.

Clone	Frequency	Mutations	K_D (nM)
PD1-1	0.05%	7	4.5
PD1-15	0.05%	10	7.3
PD1-91	0.09%	4	9.2
PD1-2	2.20%	6	9.8
PD1-7	0.05%	6	10.5
PD1-75	0.16%	5	11.1
PD1-60	0.00%	5	16.5
PD1-88	0.11%	6	27.5
PD1-4	0.09%	5	31.4
PD1-86	0.03%	4	42.6
PD1-54	0.00%	7	42.9
PD1-14	6.10%	5	45.9
PD1-82	0.07%	1	47
PD1-10	0.11%	1	47.4
PD1-11	0.38%	8	48.1
PD1-76	0.15%	1	52.6
PD1-12	0.03%	7	53.5
PD1-42	0.16%	6	53.6
PD1-43	0.59%	6	54.4
PD1-17	0.04%	7	60.1

PD1-89	0.10%	5	63.9
PD1-87	0.11%	6	63.9
PD1-48	0.12%	7	64.3
PD1-59	0.03%	5	65.9
PD1-18	0.21%	7	68.4
PD1-30	0.27%	7	68.8
PD1-23	0.14%	2	69.2
PD1-25	0.06%	6	76
PD1-8	0.13%	7	78.1
PD1-36	0.02%	6	84.9
PD1-50	0.09%	5	87.5
PD1-27	0.25%	4	88.6
PD1-21	0.96%	5	92
PD1-32	0.03%	5	92.2
PD1-56	0.08%	7	94.8
PD1-58	0.01%	6	96.7
PD1-51	0.35%	3	98.8
PD1-16	0.08%	1	100.8
PD1-77	0.08%	3	100.9
PD1-5	0.02%	7	101
PD1-84	0.09%	4	101.2
PD1-53	0.00%	7	102.8
PD1-83	0.06%	7	105.7
PD1-13	0.57%	6	106.4
PD1-3	0.07%	6	107.3
PD1-20	0.24%	6	109.4
PD1-66	0.03%	3	118.9
PD1-80	0.04%	6	125.2
PD1-63	0.08%	5	125.8
PD1-79	0.12%	6	129.2
PD1-19	0.10%	7	129.5
PD1-22	0.21%	5	132.5
PD1-81	0.04%	5	134
PD1-46	0.12%	7	136.4
PD1-26	0.10%	7	137.5
PD1-35	0.05%	6	139.1
PD1-47	0.70%	5	139.1
PD1-9	0.30%	5	148.4
PD1-55	0.07%	7	149.3
PD1-52	0.06%	1	149.4
PD1-78	0.07%	5	152
PD1-71	0.73%	7	160.9
PD1-64	0.10%	7	166.9
PD1-73	0.11%	4	170.5
PD1-67	0.03%	6	171.5

PD1-68	0.07%	6	173.2
PD1-33	0.10%	3	173.6
PD1-90	0.09%	7	174.3
PD1-72	0.18%	7	188.1
PD1-24	1.10%	6	189.8
PD1-31	0.03%	5	194.5
PD1-61	0.06%	4	194.5
PD1-37	0.13%	6	198.6
PD1-49	0.03%	7	201.4
PD1-69	0.43%	4	201.5
PD1-38	0.06%	6	203.6
PD1-57	0.76%	5	204.3
PD1-85	0.11%	6	209.6
PD1-44	0.17%	6	226.5
PD1-92	0.09%	8	235.8
PD1-62	0.09%	3	240.2
PD1-39	0.13%	4	240.5
PD1-45	0.11%	5	242.9
PD1-65	0.05%	4	260.6
PD1-28	0.02%	8	267
PD1-74	0.16%	6	285.7
PD1-95	0.05%	0	325.5

Table 5C.

			Page Supernatant (uL)		
SEQ ID NO	**CDRH3**	**0**	**10**	**40**	**160**
36	CARDATYSSGSGYW	0.06	0.39	0.42	0.33
37	CARDADYASGSGYW	0.06	0.3	0.37	0.34
38	CARDADYGSGSGYW	0.06	0.29	0.35	0.23
39	CARDFDYSSGSGYW	0.06	0.26	0.34	0.26
40	CARDADYGSGSGYW	0.06	0.3	0.3	0.3
41	CARDADYSIGSGYW	0.06	0.23	0.26	0.27
42	CARDNDYSSGSGYW	0.06	0.15	0.25	0.33
43	CARDFDYSSGSGYW	0.06	0.19	0.25	0.28
44	CARDADYGSGSGYW	0.06	0.15	0.24	0.23
45	CARDADYASGSGYW	0.06	0.14	0.23	0.32
46	CARDADYSIGSGYW	0.06	0.17	0.23	0.3
47	CARDFDYSSGSGYW	0.06	0.13	0.22	0.31
48	CARDAVYSSGSGYW	0.06	0.16	0.22	0.2
49	CARDADYASGSGYW	0.06	0.16	0.22	0.31
50	CARDWDYSSGSGYW	0.06	0.14	0.22	0.28
51	CARDFDYSSGSGYW	0.06	0.13	0.21	0.34
52	CARDADYSIGSGYW	0.06	0.16	0.21	0.23
53	CARDAFYSSGSGYW	0.06	0.16	0.21	0.28
54	CARDFDYSSGSGYW	0.06	0.13	0.21	0.34

55	CARDWDYSSGSGYW	0.06	0.12	0.2	0.16
56	CARDWDYSSGSGYW	0.06	0.13	0.19	0.26
57	CARDADYWSGSGYW	0.06	0.13	0.17	0.15
58	CARDFDYSSGSGYW	0.06	0.1	0.17	0.29
59	CARDNDYSSGSGYW	0.06	0.11	0.16	0.3
60	CARDWDYSSGSGYW	0.06	0.1	0.16	0.3
61	CARDWDYSSGSGYW	0.06	0.09	0.15	0.29
62	CARDVDYSSGSGYW	0.06	0.08	0.15	0.31
63	CARDADYKSGSGYW	0.06	0.09	0.14	0.14
64	CARDFDYSSGSGYW	0.06	0.08	0.14	0.24
65	CARDADYGSGSGYW	0.06	0.08	0.14	0.27
66	CARDALYSSGSGYW	0.06	0.09	0.14	0.2
67	CARDWDYSSGSGYW	0.06	0.08	0.13	0.26
68	CARDFDYSSGSGYW	0.06	0.08	0.13	0.25
69	CARDADYASGSGYW	0.06	0.08	0.13	0.26
70	CARDADYSSGSGYW	0.06	0.06	0.06	0.07

[00201] *Functional and developability assays*

[00202] IgGs optimized by methods described herein were tested for functional blocking of the PD-1/PD-L1 interaction. **FIG. 7A** shows that high-affinity variants demonstrated improved IC50 compared to the wild-type as well as comparator anti-PD1 antibodies having a sequence corresponding to nivolumab. IC50 and monovalent binding affinity were highly correlative. As seen in **FIG. 7B**, optimized IgGs exhibited improved binding affinity (up to 72x) and function was also increased by 9.5x. Six antibodies were identified with higher binding affinity and function than the antibody having a sequence corresponding to nivolumab. As seen in **FIG. 7C**, addition of anti-PD-1 IgGs blocked the PD-1/PDL-1 interaction, released inhibitory signal, and resulted in TCR activation and NFAT-RE-mediated luminescence (RU). Additionally, all binders retained binding to cyno PD-1 as well. Several high-affinity IgGs demonstrated low polyspecificity scores as measured by BVP binding ELISAs (**FIG. 7D**). Additionally, IgGs were tested on Unchained UNCLE machines for Tm and Tagg, as well as analytical HPLC.

[00203] 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. A nucleic acid library comprising:

a plurality of sequences comprising nucleic acids that when translated encode for antibodies or antibody fragments, wherein each of the sequences comprises a predetermined number of mutations within a CDR relative to an input sequence of an antibody;

wherein the library comprises at least 50,000 variant sequences, each represented in an amount within 1.5X of a mean frequency;

and wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 2.5X higher binding affinity than a binding infinity of the input sequence.

2. The nucleic acid library of claim 1, wherein the library comprises at least 100,000 variant sequences.

3. The nucleic acid library of claim 1, wherein at least some of the sequences encode for an antibody light chain.

4. The nucleic acid library of claim 1, wherein at least some of the sequences encode for an antibody heavy chain.

5. The nucleic acid library of claim 1, wherein each sequence of the plurality of sequences comprises at least one mutation in the CDR of a heavy chain or light chain relative to the input sequence.

6. The nucleic acid library of claim 1, wherein each sequence of the plurality of sequences comprises at least two mutations in the CDR of a heavy chain or light chain relative to the input sequence.

7. The nucleic acid library of claim 1, wherein at least one of the mutations is present in at least two individuals.

8. The nucleic acid library of claim 1, wherein at least one of the mutations is present in at least three individuals.

9. The nucleic acid library of claim 1, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 5X higher binding affinity than a binding infinity of the input sequence.

10. The nucleic acid library of claim 1, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 25X higher binding affinity than a binding infinity of the input sequence.

11. The nucleic acid library of claim 1, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 50X higher binding affinity than a binding infinity of the input sequence.

12. The nucleic acid library of claim 1,wherein each sequence of the plurality of sequences comprises at least one mutation in the CDR of a heavy chain or light chain relative to a germline sequence of the input sequence.

13. The nucleic acid library of any one of claims 1-12, wherein the CDR is a CDR1, CDR2, and CDR3 on a heavy chain.

14. The nucleic acid library of any one of claims 1-13, wherein the CDR is a CDR1, CDR2, and CDR3 on a light chain.

15. The nucleic acid library of claim 1, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having at least 70X higher binding affinity than the input sequence.

16. The nucleic acid library of claim 1, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 50 nM.

17. The nucleic acid library of claim 1, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 25 nM.

18. The nucleic acid library of claim 1, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 10 nM.

19. The nucleic acid library of claim 1, wherein the at least one sequence that when translated encodes for an antibody or antibody fragment having a K_D of less than 5 nM.

20. The nucleic acid library of claim 1, wherein the library comprises a CDR sequence of any one of SEQ ID NOs: 1-6 or 9-70.

21. The nucleic acid library of claim 1, wherein the library comprises a CDRH1, CDRH2, or CDRH3 sequence of any one of SEQ ID NOs: 1-6 or 9-70.

22. The nucleic acid library of claim 1, wherein the library comprises at least one sequence encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 10 nM.

23. The nucleic acid library of claim 1, wherein the library comprises at least one sequence encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 5 nM.

24. The nucleic acid library of claim 1, wherein the library comprises at least five sequences encoding for an antibody or antibody fragment that binds to PD-1 with a K_D of less than 10 nM.

25. The nucleic acid library of claim 1, wherein the library comprises at least 100,000 variant sequences.

26. An antibody, wherein the antibody comprises a sequence of any one of SEQ ID NOs: 1-6 or 9-70.

27. An antibody, wherein the antibody comprises a sequence of any one of SEQ ID NOs: 1-6 or 9-34; 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.

28. A method of inhibiting PD-1 activity, comprising administering the antibody of claim 26 or 27.

29. A method for treatment of a proliferative disorder, comprising administering to a subject in need thereof the antibody of claim 26 or 27.

30. The method of claim 29, wherein the proliferative disorder is cancer.

31. The method of claim 29, wherein the cancer is lung, head and neck squamous cell, colorectal, melanoma, liver, classical Hodgkin lymphoma, kidney, gastric, cervical, merkel cell, B-cell lymphoma, or bladder cancer.

32. A computerized system for antibody optimization comprising:

(a) a general purpose computer; and

(b) a computer readable medium comprising functional modules including instructions for the general purpose computer, wherein said computerized system is configured for operating in a method of:

(i) receiving operating instructions, wherein the operating instructions comprise a plurality of sequences encoding for an antibody or antibody fragment;

(ii) generating a nucleic acid library comprising the plurality of sequences comprising nucleic acids that when translated encode for antibodies or antibody fragments, wherein each of the sequences comprises a predetermined number of mutations within a CDR relative to an input sequence of an antibody; wherein the library comprises at least 50,000 variant sequences, each represented in an amount within 1.5X of a mean frequency; and wherein at least one sequence when translated encodes for an

antibody or antibody fragment having at least 2.5X higher binding affinity than a binding infinity of the input sequence; and

(iii) synthesizing the at least 50,000 variant sequences.

33. The system of claim 32, wherein the nucleic acid library comprises at least 100,000 sequences.

34. The system of claim 32, wherein the system further comprises enriching a subset of the variant sequences.

35. The system of claim 32, wherein the system further comprises expressing the antibody or antibody fragments corresponding to the variant sequences.

36. The system of claim 32, wherein the polynucleotide sequence is a murine, human, or chimeric antibody sequence.

37. The system of claim 32, wherein each sequence of the plurality of variant sequences comprises at least one mutation in a CDR of a heavy chain or light chain relative to the input sequence.

38. The system of claim 32, wherein each sequence of the plurality of variant sequences comprises at least two mutations in a CDR of a heavy chain or light chain relative to the input sequence.

39. The system of claim 32, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 5X higher binding affinity than a binding infinity of the input sequence.

40. The system of claim 32, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 25X higher binding affinity than a binding infinity of the input sequence.

41. The system of claim 32, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 50X higher binding affinity than a binding infinity of the input sequence.

42. The system of claim 32, wherein each sequence of the plurality of variant sequences comprises at least one mutation in a CDR of a heavy chain or light chain relative to a germline sequence of the input sequence.

43. The system of any one of claims 32-42, wherein the CDR is a CDR1, CDR2, and CDR3 on a heavy chain.

44. The system of any one of claims 32-43, wherein the CDR is a CDR1, CDR2, and CDR3 on a light chain.

45. The system of claim 32, wherein the antibody library has a theoretical diversity of at least 10^{12} sequences.

46. The system of claim 32, wherein the antibody library has a theoretical diversity of at least 10^{13} sequences.

47. A method for optimizing an antibody comprising:

(a) providing a plurality of polynucleotide sequences encoding for an antibody or antibody fragment;

(b) generating a nucleic acid library comprising the plurality of sequences comprising nucleic acids that when translated encode for antibodies or antibody fragments, wherein each of the sequences comprises a predetermined number of mutations within a CDR relative to an input sequence of an antibody; wherein the library comprises at least 50,000 variant sequences, each represented in an amount within 1.5X of a mean frequency; and wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 2.5X higher binding affinity than a binding infinity of the input sequence; and

(c) synthesizing the at least 50,000 variant sequences.

48. The method of claim 47, wherein the antibody library comprises at least 100,000 sequences.

49. The method of claim 47,wherein the method further comprises enriching a subset of the variant sequences.

50. The method of claim 47,wherein the method further comprises expressing the antibody or antibody fragments corresponding to the variant sequences.

51. The method of claim 47, wherein the polynucleotide sequence is a murine, human, or chimeric antibody sequence.

52. The method of claim 47, wherein each sequence of the plurality of variant sequences comprises at least one mutation in each CDR of a heavy chain or light chain, relative to the input sequence.

53. The method of claim 47, wherein each sequence of the plurality of variant sequences comprises at least two mutations in each CDR of a heavy chain or light chain relative to the input sequence.

54. The method of claim 47, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 5X higher binding affinity than a binding infinity of the input sequence.

55. The method of claim 47, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 25X higher binding affinity than a binding infinity of the input sequence.

56. The method of claim 47, wherein at least one sequence when translated encodes for an antibody or antibody fragment having at least 50X higher binding affinity than a binding infinity of the input sequence.

57. The method of claim 47, wherein each sequence comprises at least one mutation in each CDR of a heavy chain or light chain relative to a germline sequence of the input sequence.

58. The method of claim 47, wherein the nucleic acid library has a theoretical diversity of at least 10^{12} sequences.

59. The method of claim 47, wherein the nucleic acid library has a theoretical diversity of at least 10^{13} sequences.