(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 May 2019 (23.05.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/099454 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/US2018/060933

(22) International Filing Date:
14 November 2018 (14.11.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/586,800 15 November 2017 (15.11.2017) US

(72) Inventors; and
(71) Applicants: **VALADON, Philippe** [US/US]; 4650 Lisann Street, San Diego, CA 92117 (US). **ALMAGRO, Juan, Carlos** [US/US]; 320 Concord Avenue, Cambridge, MA 02138 (US).

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

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

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2019/099454 A2

(54) **Title:** HIGHLY FUNCTIONAL ANTIBODY LIBRARIES

(57) **Abstract:** The present invention relates to a method for the generation of antibody libraries of improved functionality and uses of thereof by combining a library of antibodies pre-selected for functional properties relevant to developability, such as for example improved thermal stability, with a library of CDR-H3 fragments.

HIGHLY FUNCTIONAL ANTIBODY LIBRARIES

CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This application claims benefit of priority to U.S. provisional patent application no. 62/586,800 filed on November 15, 2017, the content of which is herein incorporated by reference in its entirety.

REFERENCE TO SEQUENCE LISTING

[02] The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with the file "PCT_sequences_WorkFile.txt" created on November 14, 2018, filed on November 14, 2018 and having a size of 43 KB. The sequence listing contained in this ASCII formatted document forms part of the specification and is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[03] The present invention relates to methods for preparing antibody libraries of improved functionality and relates to the design of antibody libraries of improved functionality and uses thereof.

BACKGROUND OF THE INVENTION

[04] Antibodies are immunoglobulins, or specialized immune proteins, having a heterodimeric structure. The antibody structure consists of two heavy chains and two light chains, folded into constant and variable domains. The variable domains of the heavy chains and light chains form the antigen-binding site. Each variable region contains three hypervariable loops known as complementarity-determining regions (CDRs) which alternate with less variable regions, called framework regions (FR). Monoclonal antibodies are antibodies that are made from identical immune cells that are all clones of one parent cell. Monoclonal antibodies possess unique characteristics such as specificity, affinity, potency, stability, solubility, and clinical tolerability.

[05] Hybridoma technology, also known as monoclonal antibody technology, is an efficient means to isolate single specificity antibodies and produce them in unlimited amounts (Köhler and Milstein, 1975). This technology paved the way to generate antibodies for a diverse array of diagnostic and therapeutic applications (Reichert, 2017). In fact, antibody-based drugs represent the fastest-growing segment of all the therapeutic proteins in the biotechnology industry (Strohl, 2017).

[06] Muromonab-CD3 (Orthoclone OKT3®) was the first United States Food and Drug Administration (FDA) approved monoclonal antibody for use in therapeutic settings (Smith, 1996). The mouse monoclonal IgG2a antibody, which was developed by using hybridoma technology, blocks CD3-mediated activation of T cells and was instrumental in the prevention of organ rejection after transplantation (Emmons and Hunsicker, 1987). However, patients who were treated with Orthoclone OKT3® developed a significant percentage of anti-drug antibodies (ADA). The development of ADA is also known as a "human anti-mouse antibody" (HAMA) response (Kimball et al., 1993). The HAMA response leads to the inactivation and elimination of murine antibodies (Shawler et al., 1985). Further, the HAMA response prevents the use of multiple administrations of an antibody, for example required for cancer therapy. Further complicating the use of murine monoclonal antibodies in human therapy is their association with the generation of severe allergic reactions. Together, such issues hamper the use of murine antibodies in human therapy (Shawler et al., 1985).

[07] To engineer more human-like antibodies and, thus, increase efficacy while decreasing immunogenicity, nonhuman variable (V) domains are combined with human constant (C) domains to generate molecules with 70% or more human content. This method is called chimerization and led to the approval of the first cancer-treating therapeutic antibody, Rituximab (Rituxan ®) (Morrison et al., 1984). Rituximab has been a tremendous medical and commercial success, currently being the fourth best-selling innovative drug of any kind (Strohl, 2017).

[08] Additional technology platforms emerged in parallel with Rituximab, which aimed to generate more human-like antibodies. Such technology platforms were perfected during the last three decades and include humanization (Jones et al., 1986), selection of fully human antibodies from Fv and Fab phage-displayed libraries (McCafferty et al., 1990) and the development of transgenic animals capable of generating fully human antibodies (Green et al., 1994; Lonberg et al., 1994).

[09] Humanized antibodies have more human content than chimeric antibodies but still do not eliminate the possibility of human anti-human antibody (HAHA) responses (Nechansky, 2010). Transgenic mice are capable of producing fully human antibodies. However, immunization does not always result in a successful *in vivo* antibody response with the desired level of affinity (Green and Jakobovits, 1998; Pruzina et al., 2011). This is particularly true for conserved epitopes between human and mouse orthologs. Transgenic rats and more recently, transgenic OmniChicken, may partially mitigate this limitation (Ma et al., 2013 and Ching et al., 2017). However, toxic and unstable antigens and proteins with allosteric conformational changes are not well suited for an immunization approach and require an alternative solution for antibody discovery.

[10] Phage display technology connects proteins displayed at bacteriophage surface to their genes in the bacteriophage's genome. Thus, phage display opened the possibility of designing and manipulating

the repertoire of antibody genes to be used as source of antibodies in phage antibody libraries, hence enabling the selection of fully human antibodies (Hanes and Pluckthun, 1997; Francisco et al., 1993; Beerli et al., 2008; Cherf and Cochran, 2015; Finlay and Almagro, 2012). Moreover, these technology platforms allow selection for desired pre-defined epitopes, thus avoiding immunodominant epitopes. Further, these technology platforms can also focus on specific conformations, rare cross-reactive epitopes, or conditions to select antibody variants with enhanced biophysical and biochemical properties.

[11] The concept that more diverse, functional, and larger libraries produce a larger number of specific antibodies with higher affinity antibodies is largely expected. One mathematic model has been proposed to formalize such a concept (Perelson and Oster, 1979). In this model, the probability (P) that an epitope is recognized by at least one antibody in a repertoire depends on the probability (p) that an antibody recognizes a random epitope with an affinity above a threshold value and on the size of the library (N) according to the equation P = 1-e^(NP). This model predicts that the larger the size of the functional repertoire, the higher the chances of finding a specific and higher affinity antibody.

[12] In practice, it has been shown that a human antibody fragment phage display library of 10^{10} members yields ~90 nM antibody fragments to four different proteins (Marks et al., 1991). Larger libraries, e.g., > 10^{10} (Vaughan et al., 1996; de Haard et al., 1999), have produced antibody fragments with single-digit nM or sub-nM affinities. Similarly, a library called Griffiths' library (Griffiths et al., 1994) of > 10^{10} members produced sub-nM binders, whereas, antibody fragments of > 100 nM were obtained when a small portion of the library containing 10^{7} clones was used in the selection process. Together, these results support the concept that larger libraries generate higher affinity antibodies.

[13] The theoretical number of unique antibody variants is virtually infinite. For instance, if the six CDRs that form the antigen-binding site were diversified with the 20 natural amino acids and it is considered that each CDR has on average seven positions to diversify, the corresponding theoretical number of unique variants is $20^{(6 \times 7)} = 10^{54}$. However, only a very small fraction, which does not exceed 10^{11} unique variants, can be cloned and displayed on the phage surface (Hoogenboom, 2005). More importantly, these 10^{11} unique antibody variants should be functional to maximize the probability of producing antibodies with reasonable specificity and affinity, roughly in the low nM range.

[14] Several strategies have been described in the art to maximize the number of functional antibody variants in a phage display library. One typical strategy consists in cloning human antibody genes from natural sources, such as peripheral blood mononuclear cells (PBMCs) (de Haard et al., 1999). Other strategies are the construction of fully synthetic libraries, which are computationally designed and generated by chemical synthesis (Griffiths et al., 1994; Knappik et al., 2000; Shi et al., 2010), and semisynthetic libraries, which combine natural diversity and synthetic diversity (Hoet et al., 2005).

[15] Antibody libraries composed of natural diversity, also known in the art as naïve libraries, were the first generation of phage displayed antibody libraries (Marks et al., 1991). Although successful, naïve libraries included antibody genes toxic to *E. coli* and thus with low expression levels or no expression at all on the phage surface. These toxic genes severely compromised the number of functional antibodies displayed in the library. Synthetic antibody libraries followed and partially mitigated this limitation (Knappik et al., 2000) (US patent 6,828,422). However, synthetic libraries must carefully be designed by making assumptions regarding the number of positions to diversify and type of amino acids to include in the design and proportion of each amino acid per position. These assumptions do not always hold true, particularly for CDR-H3, which is by far the most diverse region of the antigen-binding site and key in defining the specificity and affinity of antibodies. The structure of CDR-H1 and CDR-H2 can be predicted with an accuracy of < 1.0 Å. This is a critical step in the successful design of the diversity of a synthetic library. However, no current method is available in the art to reliably predict the CDR-H3 structure (Almagro et al. 2014).

[16] The quality of the synthesis process also impacts the functionality of the library. Nucleotide sequences with stop codons leading to truncated sequences do not produce functional antibody fragments fused to a virion particle. Insertions or deletions of one or two nucleotides changing the reading frame of the antibody gene can generate stretches of amino acids that may impair folding or produce clones with hydrophobic amino acids resulting in aggregation. These non-functional antibody variants compromise the functionality of the library leading to poorly performing libraries and, overall, lower and often no production of antibodies with the desired specificity and affinity originally sought.

[17] The diversification methods used to generate synthetic libraries greatly influence the quality of the output antibodies. For instance, three libraries were designed for affinity maturation of anti-Oncostatin (anti-OSM) antibodies (US patent 8,580,714). The diversity of libraries was designed at positions frequently observed in contact with protein and peptide antigens (Raghunathan et al., 2012). The diversification regime, or amino acids and their frequencies per diversified position, were different in the three libraries. Two libraries were designed with few amino acids found in known antibody sequences. The third library was generated with random (NNK) codons, which produce the 20 amino acids plus one stop codon. However, the libraries designed with amino acids found in known antibody sequences yielded more diverse and higher affinity antibodies.

[18] Further, it has been realized that antibodies selected from naïve, synthetic, and semisynthetic libraries often tend to fail during the development stages of formulation and manufacturing due to sub-optimal biophysical properties. This occurs despite possessing the desired specificity and affinity. Antibodies undergo posttranslational modifications (PTM) of some amino acids such as deamidation of asparagine (N), oxidation of methionine (M), and isomerization of aspartic acid (D) (Gilliland et al.,

2012). These chemical modifications of the amino acids can result in heterogeneities in the antibody preparation and/or lack of potency if the amino acids are involved in the interaction with the antigen. In other instances, exposure of tryptophan (W) to the solvent can induce aggregation, thus leading to immunogenic reactions or lack of solubility at concentrations required for the therapeutic indication (Bethea et al., 2012).

[19] Therefore, there has been a continued need in the art for making functional antibody libraries that produce antibodies amenable to clinical development and hence increase the probability of success along the preclinical and manufacturing processes (Urlinger et al., US Patent 9,541,559). Computational methods and design principles, collectively called developability predictive methods (Kumar and Sigh, Developability of Biotherapeutics: Computational Approaches. CRC Press. 2015), have been perfected in the art to identify and remove residues that pose developability liabilities. It is critical to experimentally assess antibody variants as early as possible during the antibody discovery process. Developability predictive methods significantly improve the chances of successful antibody development, manufacturing, formulation, and stabilization to achieve the desire therapeutic effect.

[20] This invention discloses highly functional antibody libraries in addition to methods to make such highly functional antibody libraries. The exemplary antibody libraries of this invention combine highly stable and developable antibody variants with natural diversity in CDR-H3, as well as methods to remove non-productive antibody sequences such as out-of-frame variants and/or poorly folded antibodies due to imprecisions in the design, without compromising the functional size of the library.

SUMMARY OF THE INVENTION

[21] The present invention includes methods for producing highly functional antibody libraries, comprising the steps of: (1) designing and preparing primary antibody libraries (PLs); (2) applying a selection process (filtration) to enrich the PLs of step 1 with variants of improved functionality and/or developability to prepare intermediate filtered libraries (FLs); and (3) preparing highly functional secondary antibody libraries (SLs) by combining FLs with diverse CDR-H3 fragments.

[22] Two PLs were designed so that each of the PLs had one V_H scaffold binding to a *Staphylococcus aureus* Protein A. Two distinct V_L scaffolds were designed as counterpart of the V_H scaffold. One of the V_L scaffolds contained a short CDR-L1 loop (PL1). The other V_L scaffold contained a long CDR-L1 loop (PL2) (**Figure 1**). By changing the length of CDR-L1 from a short to a long loop, antibodies alter the preference to bind protein or peptide targets, respectively (Vargas-Madrazo et al., 1995; Raghunathan et

al., 2012). Therefore, by using the proper $_L$Vscaffold, the libraries of this invention can be used for selection of antibodies against protein or peptide targets. Selection of the two libraries, PL1 and PL2, would potentially generate antibodies that bind diverse types of epitopes on a given target.

[23] The scaffolds of the two PLs were diversified in positions that bind both protein and peptide targets and amino acids observed in human germline genes and natural antibodies. Amino acids associated with developability liabilities were avoided. Such amino acids included: (i) asparagine (N) followed by any amino acid but proline (XnoP) followed by serine (S) or threonine (S/T) [NXnoP(S/T)], which generates N-glycosylation sites; (ii) aspartic acid (D) followed by glycine (G) [DG], which tends to isomerize; (iii) asparagine (N) followed by glycine (G) or serine [NG/S], which tends to deamidate, (iv) exposed methionine (M), which tends to oxidize and (v) exposed tryptophan (W), which leads to aggregation spots.

[24] In one aspect of the invention, it was reasoned that by using only one CDR-H3 sequence to generate the PLs, the diversity of amino acids in contact with, or nearby, said CDR-H3 may be constrained to a few and specific residues to accommodate said CDR-H3 under specific selection conditions. Therefore, a set of CDR-H3 sequences called "neutral H3Js" was designed by starting from the repertoire of human D genes (IGDH) and combining them with the human heavy chain J genes (IGJH) in the germline configuration. Since antibodies from natural primary repertoires are mostly sequences in germline configuration, it was hypothesized that the set of neutral H3Js had enough diversity to avoid biases in amino acids at V_H and V_L in contact with, or nearby, the neutral H3J sequences, thus providing a diverse and favorable environment to select for developable diversified scaffolds and support cloning of natural CDR-H3 fragments in the third step described below.

[25] The designed V_H and V_L scaffolds were assembled as single chain Fv fragments (scFv) in a V_L-linker-V_H configuration and synthesized using trinucleotide phosphoramidites. Trinucleotide phosphoramidites synthesis, also known as trimer technology, is a type of synthesis that is based on synthetic codons instead of single nucleotides. Trinucleotide phosphoramidites synthesis generates precise combinations of amino acids at specifically targeted positions for diversifications while avoiding stop codons and unwanted amino acids which may disrupt the folding of the scaffolds used to generate the libraries. The quality control of the synthetic fragments was assessed *via* sequencing of 96 fragments in each library. The results indicated that 50% to 60% sequences were in-frame and matched the design.

[26] The synthetic fragments were then cloned in a phage display vector and displayed on the surface of M13 phage as fusion proteins to its minor coat protein pIII to generate PL1 and PL2 following standard molecular biology protocols well known to those skilled in the art and described herein. Display on other platforms including yeast or related display technologies is a clear extension of the invention. A sample of individual clones chosen at random from PL1 and PL2 were submitted to Sanger sequencing. The

percentage of in-frame sequences matching the design was 61.0% and 87.3%, for PL1 and PL2, respectively, which was in close agreement with the quality of the synthetic fragments.

[27] In a second aspect of the invention, there is provided the frequency of Protein A binders among the clones from the PLs displayed as scFvs after diverse incubation times and temperatures. It has been established that the least stable domain of the human IgG1 is the C_H2 (Gilliland et al., 2012), which unfolds at 68°C. The question of whether a significant number of clones from the PLs were still stable at 68°C or above and hence, yielded developable antibodies when expressed as IgG1 in a therapeutic antibody was addressed. The percentage of clones binding Protein A after incubation at 70°C for 10 min was 48.8% and 55.8% for PL1 and PL2, respectively, down from 58.5% and 69.8%, respectively in the PL1 and PL2. The difference of 9.7% and 14.0% less clones binding Protein A in PL1 and PL2, respectively, after heat shock indicated that some variants were unstable at 70°C.

[28] In a third aspect of the invention, PL1 and PL2 were incubated for 10 min at 70°C and well-folded and developable antibody fragments were rescued with Protein A, while unstable and non-developable antibody variants, which were either denatured or aggregated, were removed by a simple washing. Other harsh conditions include, but by no means are limited to, other temperatures and incubation times, high or low pH, high salt concentrations, and protease digestion. In one preferred embodiment, there exist counter-selections to remove antibodies with tendencies to aggregate, for example from the interactions of exposed hydrophobic residues.

[29] After incubation at 70°C for 10 min of PL1 and PL2 and rescue of well-folded antibody variants with Protein A, clones chosen at random from the filtered libraries (FLs) were sequenced and assayed for Protein A binding. The percentage of unique scFvs binding Protein A was around 90% in both libraries, 89.5% and 89.2%, respectively in FL1 and FL2, with virtually all the sequences being in-frame, 94.7% and 97.3%, respectively. Therefore, the functionality measured as the ability to bind Protein A was significantly improved by 31.0% and 19.4%, respectively in FL1 and FL2, after the heat-shock and filtering with Protein A.

[30] It should be noted that other natural ligands binding well-folded antibodies outside the antigen-binding site have been described in the art. For example, *Peptostreptococcus magnus* Protein-L binds the V_L scaffold with the long CDR-L1 loop. The use of Protein-L as a ligand to select well-folded antibodies, alone and/or in conjunction with Protein A is a clear extension of this invention.

[31] Natural ligands that bind variable regions of antibodies, such as Protein A and Protein L, have extensively been used in the prior art to select for stable antibody domains after incubation under denaturing or destabilizing conditions. However, their use has been limited to enrich the final antibody library and/or select for variants after mutagenesis to improve stability. As shown herein, the selection with Protein A after submitting the libraries for harsh incubation conditions, removes a significant

number of variants from the library. The decrease in the number of variants, and thus of the library size, depends on the library quality, scaffolds used to build the library, design of the library diversity, and selection conditions for stable and functional antibody variants. In the exemplary proofs of this invention, incubation at 70°C for 10 min led to a reduction of well-folded variants in PL1 and PL2 of 9.7% and 14.0% respectively. PL1 and PL2 are built with two human V_L scaffolds that belong to two different germline gene families and only share 68.3% identities. The kinetic of unfolding, as showed in the examples of this invention, is significantly different. However, in both cases, a consistent average decrease of ~12% in Protein A binders represents a loss of 1.2×10^9 unique and potentially functional antibody variants in a library containing 1×10^{10} unique antibody variants. Therefore, improving stability after filtering came with the price of reducing the diversity of the libraries.

[32] In a forth aspect of the present invention, there is provided a method designed to restore the diversity of the library. In this step, the nucleotide sequences encoding the FLs were amplified by molecular biology techniques known to those skilled in the art and combined with diverse natural CDR-H3 fragments, called "natural H3Js" to produce the SLs. It was reasoned that by replacing the neutral H3Js fragments in the FLs with natural H3Js, highly stable and functional libraries can be obtained. This rationale is supported by a substantial body of work indicating that CDR-H3 is key to determine the specificity and affinity of the antibodies and hence antibody libraries with highly diverse CDR-H3 fragments should increase the probability of obtaining diverse and high affinity antibodies. In the present invention, the natural H3Js were isolated from a pool of 200 donors by molecular biology methods known to those skilled in the art. Amplification of diverse CDR-H3 fragments from other sources including CDR-H3 fragments obtained by synthetic means are clear extensions of this invention.

[33] Analysis of a sample of antibody variants chosen at random indicated that 68.3% and 75.6% in SL1 and SL2, respectively, bound Protein A when expressed as scFvs on the phage surface after incubation at 70°C for 10 min. Moreover, the sequence of all these clones had natural and unique CDR-H3 sequences. Comparatively to the PLs, this represents an increase of 19.5% and 19.8% in ability to survive a heat shock in SL1 and SL2, respectively. In relative numbers, these values translate in an average 36% to 40% increase of the stability of the SLs versus the PLs after incubation at 70°C for 10 min.

[34] It should be emphasized that it was not obvious from prior art that by combining the highly stable variants comprising the FLs with highly diverse natural CDR-H3 fragments, obtained from natural sources, which have not been under selection on harsh conditions such as incubation at 70°C for 10 min, the resulting SLs would retain the property of being highly stable. The selection of libraries of antibody fragments and antibody domains by Protein A had a profound effect on the diversity of the CDRs. Therefore, the gain in stability was at the expense of the library diversity. In the present invention, the

combination of highly stable scaffolds used to build the libraries, a design based on positions found in contact with antigens and solvent exposed, germline gene diversity to diversify those positions, removal of liability developabilities, and the selection of PLs in conjunction with the set of neutral H3Js, led to a collection of highly stable variants suited to accommodate a collection of highly diverse natural H3J fragments, and hence, a highly functional antibody library.

[35] For each step of the construction, primary, filtrated and secondary libraries were analyzed by next generation sequencing (NGS) and established the conformity of the primary libraries to the intended design, a bias toward more hydrophilic sequences around CDR-H3 during the filtration process compensated by a diverse environment offered by the set of neutral HJ3 fragment, and the very high diversity of the secondary libraries. Analysis of Protein A binding from single clones provided the proof-of-concept that the overall filtration process resulted in stability improvement at the library level, while NGS analysis confirmed the retention of the in-frame character of the vast majority of the clones during the construction of the secondary libraries together with a boost in diversity.

[36] Finally, to demonstrate the potential of the SLs to produce specific and high affinity antibodies, pannings with two known antigen models have been performed. In one exemplary proof of this invention, the antigen model is Tumor Necrosis Factor (TNF). In another example, the antigen model is human serum albumin (HSA). After four rounds of selection against TNF and three for HSA, diverse and specific antibodies were obtained.

BRIEF DESCRIPTION OF THE DRAWINGS

[37] Those skilled in the art will recognize that the drawings described below are for illustrative purposes only. The drawings are not intended to limit the scope of the invention but to provide exemplary embodiments.

[38] **Figure 1** provides a ribbon representation of the V_H and V_L scaffolds. **Figure 1A** shows a drawing showing the ribbon representation of an Fv with a short L1 loop (PDB ID: 1ILC) [3-20/3-23]. **Figure 1B** shows a drawing showing the ribbon representation of an Fv with a long CDR-L1 loop (PBD ID: 1ILD) [4-01/3-23]. These drawings show the ribbon representations of the Fvs used to build the semi-synthetic libraries of this invention. These Fvs have been solved by x-ray crystallography in a Fab configuration in association the CDR-H3 loop of a known therapeutic antibody (CNTO 888; (Obmolova et al., 2012).

[39] **Figure 2** provides a list of PDB ID complex structures used to design the primary repertoires. This dataset of unique antibody:antigen complexes was obtained by starting from all the antibody structures compiled at the PDB and curated by the IMGT as of March 17th, 2017. The initial dataset consisted of 2,645 antibody structures from diverse species and specificities. This initial data set was mined to extract the unique and well-solved antigen:antibody listed in the table.

[40] **Figure 3** provides a table showing a diversification regime at CDR-H1 and CDR-H2 of the V_H scaffold. The residues in contact with antigens were determined in the set of structures listed in **Figure 2** and mapped onto the structures of the V_H scaffold paired with the two V_L scaffolds PDB IDs: 1ILC and 1ILD. A total of 10 positions were targeted for diversification, four in the CDR-H1 and six in the CDR-H2. The diversification regime was designed using three sources of information: the dataset of curated antibody structures listed in **Figure 2**, the V_H sequences available at NCBI, and the germline genes of the human IGHV3 family compiled at the IMGT. The estimated number of amino acids per position is listed in the last column of the figure and yields a diversity of 5.4×10^6 unique amino acid V_H sequences.

[41] **Figure 4** shows the diversification regime of the 3-20 V_L scaffold. To identify positions for diversification and the diversification regime the same procedure than that for the V_H scaffold was followed. The estimated number of amino acids per position is listed in the last column of the figure and yields a diversity of 1.1×10^6 unique amino acid V_L sequences.

[42] **Figure 5** shows the diversification regime of the 4-01 V_L scaffold. To identify positions for diversification and the diversification regime the same procedure than that for the V_H scaffold was followed. The estimated number of amino acids per position is listed in the last column of the figure and yields a diversity of 1.3×10^6 unique amino acid V_L sequences.

[43] **Figure 6** provides the configuration of the two primary scFv repertoires, diversified regions and cloning sites. V_L and V_H are linked to form a scFv by the repetitive stretch of amino acids GS19. The V_L-linker-V_H configuration places the H3J fragment on the C-terminal side. Two BglI/SfiI sites located on each side of the construct allow for cloning into the acceptor vector. NcoI and KpnI sites are common to both repertoires, same for the JK1 anchor in the light chain sequence and the human VH Conserved Motif (CM) in the heavy chain sequence.

[44] **Figure 7** provides a graph showing phage binding to Protein A by ELISA. The y-axis shows the optical density measurements at 490 nm and the x-axis shows dilutions in virions/ml for the 3-20/3-23 control scFv, the 4-1/3-23 control scFv, the 3-20/3-23 library, and the 4-01/3-23 library. COSTAR plates 3369 (Corning) were coated with Protein A (Sigma Aldrich, cat# P6031) at 4 μg/ml in TBS overnight at 4°C. After blocking with TBS with 0.1% Tween® 20 v/v (TBST) and 5% w/v nonfat dry milk for one hour, virions (~2.6×10^{12} virions/ml) in TBST with 5% w/v nonfat dry milk were added to the wells, 2-fold serially diluted and incubated for 2 h at 37°C. As a reference, virions derived from the parent

scaffolds with the CDR-H3 of CNTO 888 cloned in the same vector were added and similarly diluted on the plate. Bound phage was detected with A4G1.6 monoclonal antibody (Antibody Design Labs, San Diego, CA) conjugated to HRP. Binding of the secondary antibody, a murine IgG1, to Protein A was blocked by polyclonal human IgG at 100 μg/ml added to the incubation buffer.

[45] **Figure 8** provides a graph showing, on the y-axis, optical density measurements at 490 nm taken at different temperature points during heat shock showing the thermal unfolding of the 4-01/3-23 and 3-20/3-23 control scFvs. The 3-20/3-23 and 4-01/3-23 control scFvs displayed as fusion proteins to pIII on the phage surface were incubated for 10 min over a range of temperatures, starting at 40°C and increasing the temperature up to 80°C in steps of 5°C. The unfolding process is monitored with a direct Protein A ELISA shown in **Figure 7**.

[46] **Figure 9** provides a graph showing, on the y-axis, optical density measurements at 490 nm taken at different time points during heat shock showing the thermal unfolding of the 4-01/3-23 and 3-20/3-23 control scFvs at 60°C and 72°C. The 3-20/3-23 scFv and 4-01/3-23 scFv were displayed as fusion proteins to pIII on the phage surface and the unfolding process is monitored with a direct Protein A ELISA shown in **Figure 7**.

[47] **Figure 10** provides a bar graph showing different Protein A binding relative to control scFvs at 37°C. The binding shown is measuring PL1 single clone scFv-phages binding to Protein A relative to control scFv. A sample of 44 ampicillin-resistant colonies was chosen at random from PL1. The phage displaying scFvs were prepared in 96 deep-well plates and, after pelleting the bacteria by centrifugation, 100 μl aliquots were incubated for 10 min at 37°C in a 96-well PCR plate. After cooling down the samples for 30 min on ice, binding to Protein A shown in **Figure 7** was performed. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[48] **Figure 11** provides a bar graph showing different Protein A binding relative to control scFvs at 37°C. The binding shown is measuring PL2 single clone scFv-phages binding to Protein A relative to control scFv. A sample of 47 ampicillin-resistant colonies was selected from the PL2. Three clones, either with the parent vector or a partial insert were not further studied. The resulting 44 clones were analyzed as described in **Figure 10**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[49] **Figure 12** provides a bar graph showing residual Protein A binding of PL1 scFv-phage single clones following 10 min at 70°C. The binding shown is measuring PL1 single clone scFv-phages binding to Protein A relative to control scFv. The 44 clones selected from PL1 were incubated for 10 min at 70°C as previously described in **Figure 10**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[50] **Figure 13** provides a bar graph showing residual Protein A binding of PL2 scFv-phage single clones following 10 min at 70°C. The binding shown is measuring PL2 single clone scFv-phages binding to Protein A relative to control scFv. The 44 clones selected from PL2 were incubated for 10 min at 70°C as previously described in **Figure 10**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[51] **Figure 14** provides a bar graph showing Protein A binding relative to control scFvs after incubation at 70°C for 10 min and Protein A filtering. The binding shown is measuring FL1 single clone scFv-phages binding to Protein A relative to control scFv. A sample of 44 ampicillin-resistant colonies was chosen at random from FL1. Phage were prepared and assayed on Protein A as described in **Figure 10**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[52] **Figure 15** provides a bar graph showing Protein A binding relative to control scFvs after incubation at 70°C for 10 min and Protein A filtering. The binding shown is measuring FL2 single clone scFv-phage binding to Protein A relative to control scFv. A sample of 44 ampicillin-resistant colonies was selected at random from FL2. The phages were prepared and assayed on Protein A as described in **Figure 10**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[53] **Figure 16** provides a drawing depicting a strategy for assembling seamlessly semisynthetic secondary repertoires. The top line shows a strategy for the filtrated primary library scaffold amplification, the middle line shows a strategy for an assembly with natural H3J fragments, and the bottom line of **Figure 16** shows the final product. The diversified scaffold fragments from the filtrated primary libraries were amplified by a pair of primer located in the pelB leader and 5' of the human VH Consensus Motif (CM). Two BsaI sites added at the end of the C-terminal of the Fv and at the beginning of the natural H3J fragments are used to assemble the complete scFv in a single digestion step plus ligation reaction. Finally, the assembled product is amplified as a whole scFv prior to ligation and cloning.

[54] **Figure 17** shows a series of photographs of DNA gel electrophoresis, comparing between SL1 (on the left) and SL2 (on the right) library assembly by seamless amplification. **Figure 17A** shows an electrophoresis gel of the amplified filtrated primary library scaffolds. **Figure 17B** shows an electrophoresis gel before (left lane) and after (right lane) the fragment seamless assembly with natural H3J fragments by simultaneous digestion by BsaI and ligation. **Figure 17C** shows an electrophoresis gel of the amplification of the full scFv fragments.

[55] **Figure 18** provides a bar graph showing Protein A binding of SL1 single clones relative to control scFvs after incubation at 37°C for 10 min. A sample of 44 ampicillin-resistant colonies was selected at random from the SL1. The phage were prepared and assayed on Protein A as described in **Figure 7**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[56] **Figure 19** provides a bar graph showing Protein A binding of SL2 single clones relative to control scFvs after incubation at 37°C for 10 min. A sample of 44 ampicillin-resistant colonies was selected at random from the SL2. The phages were prepared and assayed on Protein A as described in **Figure 7**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[57] **Figure 20** provides a bar graph showing Protein A binding of SL1 single clones relative to control scFvs after incubation at 70°C for 10 min. The phages were prepared and assayed on Protein A as described in **Figure 7**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[58] **Figure 21** provides a bar graph showing Protein A binding of SL2 single clones relative to control scFvs after incubation at 70°C for 10 min. The phages were prepared and assayed on Protein A as described in **Figure 7**. In-frame clones are indicated by the sign '=' and double transformants by the sign '#'.

[59] **Figure 22** provides a bar graph showing a comparison of Protein A survival between PLs and SLs after incubation at 70°C for 10 min. Survival is defined ≥10% with respect to the control scaffold in the Protein A binding ELISA after incubation for 10 min at 70°C for phage clones having 25% or more binding to Protein A relative to the control scaffolds.

[60] **Figure 23** provides a graph showing, on the y-axis, optical density measurements at 450 nm versus 570 nm showing the binding of the purified anti-TNFalpha scFv TNF-E12 taken at different concentrations on the x-axis to either TNFalpha or BSA as a negative control.

DEFINITIONS

[61] Detailed descriptions of preferred embodiments are provided herein. It is to be understood, however, that the present invention may be embodied in various forms. Therefore, specific references to various forms are provided as a basis for the claims and for teaching one skilled in the present art to employ the present invention in appropriate system, structure, or manner.

[62] 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 to which this invention belongs. All patents, applications, published applications, and other publications referred to herein are incorporated by reference in their entirety. If a definition set forth in this section is contrary to or otherwise inconsistent with a definition set forth in the patents, applications, published applications, and other publications that

are herein incorporated by reference, the definition set forth in this section prevails over the definition that is incorporated herein by reference.

[63] For the purposes of this specification, the following definitions will apply and whenever appropriate, terms used in the singular will also include the plural and vice versa. In the event that any definition set forth below conflicts with the usage of that word in any other document, including any document incorporated herein by reference, the definition set forth below shall always control for purposes of interpreting this specification and its associated claims unless a contrary meaning is clearly intended (for example in the document where the term is originally used). It is noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the," include plural referents unless expressly and unequivocally limited to one referent. The use of "or" means "and/or" unless stated otherwise. For illustration purposes, but not as a limitation, "X and/or Y" can mean "X" or "Y" or "X and Y". The use of "comprise," "comprises," "comprising," "include," "includes," and "including" are interchangeable and not intended to be limiting. Furthermore, where the description of one or more embodiments uses the term "comprising," those skilled in the art would understand that, in some specific instances, the embodiment or embodiments can be alternatively described using the language "consisting essentially of" and/or "consisting of". The term "and/or" means one or all of the listed elements or a combination of any two or more of the listed elements.

[64] The section headings used herein are for organizational purposes only and are not to be construed as limiting the described subject matter in any way. All literature cited in this specification, including but not limited to, patents, patent applications, articles, books, and treatises are expressly incorporated by reference in their entirety for any purpose. In the event that any of the incorporated literature contradicts any term defined herein, this specification controls. While the present teachings are described in conjunction with various embodiments, it is not intended that the present teachings be limited to such embodiments. On the contrary, the present teachings encompass various alternatives, modifications, and equivalents, as will be appreciated by those of skill in the art.

[65] The practice of the present invention may employ conventional techniques and descriptions of bacteriology, molecular biology (including recombinant techniques), cell biology, and biochemistry, which are within the skill of the art. Such conventional techniques include PCR, extension reaction, oligonucleotide synthesis and oligonucleotide annealing, ELISA. Specific illustrations of suitable techniques can be added by reference to the example herein below. However, other equivalent conventional procedures can, of course, also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals such as Genome Analysis: A Laboratory Manual Series (Vols. I-IV), PCR Primer: A Laboratory Manual, and Molecular Cloning: A Laboratory Manual (all from Cold Spring Harbor Laboratory Press, 1989), Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984,

IRL Press, London, Nelson and Cox (2000), Lehninger, Principles of Biochemistry 3rd Ed., W. H. Freeman Pub., New York, N.Y. and Berg *et al.* (2002) Biochemistry, 5th Ed., W. H. Freeman Pub., New York, N.Y. Barbas CF *et al.*, (2004) Phage Display: A Laboratory Manual. CSHL Press, all of which are herein incorporated in their entirety by reference for all purposes.

[66] The term "antibody", as used herein, is used in the broadest sense and refers to monoclonal antibodies and one or more fragments of an antibody that retain the ability to specifically bind to an antigen (e.g., lysozyme). It has been shown that the antigen-binding function of an antibody can be performed by fragments of a full-length antibody. Examples of antigen-binding fragments encompassed a Fv fragment consisting of the V_L and V_H domains of a single arm of an antibody, a Fab fragment, a F(ab)2, which is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region fragment a monovalent fragment consisting of the V_L, V_H, CL and C_H1 domains, a Fd fragment consisting of the V_H and C_H1 domains.

[67] Furthermore, although the two domains of the Fv fragment, V_L and V_H domains are coded for by separate genes, they can be joined, using recombinant methods, by a synthetic linker that enables them to be made as a single protein chain in which the V_L and V_H regions pair to form monovalent molecules, known as single chain Fv (scFv). Such scFvs are also intended to be encompassed within the term "antibody". These antibody fragments are obtained using conventional techniques known to those of skill in the art and the fragments are screened using, but not limited to, phage, yeast and mammalian display for utility in the same manner as are intact antibodies.

[68] As used herein, "antibody variable domain" refers to the portions of the light and heavy chains of antibody molecules that interact specifically with an antigen (e.g., lysozyme). "V_H" or "V_H domain" refers to a variable domain of an antibody heavy chain. "V_L" or "V_L domain" refers to a variable domain of an antibody light chain. The V_L domain is produced by the recombination of the IGVL and IGVJ germline genes, whereas the V_H domain is encoded by repertoires of IGVH, IGVD and IGJH germline genes. V_H and V_L contain the antigen-binding site and hence define the capacity of antibodies to bind virtually any antigen with exquisite specificity and high affinity.

[69] The term "antigen-binding site", as used herein, contains the portion of the variable domains that interact with the antigens. Definitions of the antigen-binding site include the complementarity-determining regions (CDRs) as defined by Kabat (Kabat *et al.*, Sequences of Proteins of Immunological Interest, 5th ed. Bethesda, Md.: National Center for Biotechnology Information, National Library of Medicine, 1991). There are three CDRs in V_L: CDR-L1, CDR-L2, and CDR-L3, and three in V_H: CDR-H1, CDR-H2, and CDR-H3. The CDRs alternate with conserved regions called Framework Regions (FRs), four in V_L: FR-L1, FR-L2, FR-L3 and FR-L4, and four in V_H: FR-H1, FR-H2, FR-H3 and FR-H4. The antigen-binding site can also be defined as specificity-determining regions as defined and the SDR

usage regions (SDRUs). There are six SDRs or SDRUs, which approximately correspond with three CDRs - a comparison of the different definitions of antigen-binding site can be found at Gilliland *et al.*, Methods Mol Biol. 841:321, 2012.

[70] As used herein, "amino acid position" refers to a position of an amino acid located in V_H or V_L amino acid sequence. Several numbering systems have been proposed to identify an amino acid position. In this invention, the Chothia's numbering convention (Chothia and Lesk, 1987) is used.

[71] A "diversified position" refers herein to an amino acid position with different amino acids represented at said position. In one aspect of the invention, positions to diversify are determined by identifying amino acids in the antigen-binding site contact with antigens in the structure of antibody:antigen complexes determined by x-ray crystallography. Antibody amino acids in contact with the antigen are defined by the distance between an atom of said amino acid in the antibody and an atom of an amino acid in the antigen. Typically, two atoms are in contact when distance between said atoms is < 3.5 Å. A compilation of amino acids in antibodies in contact with antigens in known antigen:antibody structures is compiled at the IMGT. In other aspects of the invention, the positions to diversify are defined by the exposure of the amino acids to the solvent, defined as accessible surface area (ASA).

[72] The "diversification regime" as used in this invention refers to amino acids used to diversify the amino acid position. The diversification regime is derived from amino acid sequences of known and/or naturally occurring antibodies or antigen binding fragments. Diversified positions are typically found in the CDRs in known and/or naturally occurring antibodies and their discovery are facilitated by the antibody sequences and structures compiled at Internet-based databases. Databases compiling amino acid structures include the Protein Data Bank (PDB; http://www.rcsb.org/pdb/). Antibody sequence databases include V-base (http://www2.mrc-lmb.cam.ac.uk/vbase/), The National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/), the IMGT, and Abysis (http://www.bioinf.org.uk/abs/). These electronic resources provide extensive collections and alignments of human light and heavy chain sequences and facilitate the determination of highly diverse positions in these sequences.

[73] As used herein, "repertoire" or "library" refers to a plurality of antibodies, antibody fragment sequences, antibody variable domains, diversified scaffolds, or the nucleic acids that encode these sequences, the sequences being different in the combination of variant amino acids that are introduced into these sequences according to the methods of the invention.

[74] A "scaffold", as used herein, refers to a polypeptide or portion thereof that maintains a stable structure or structural element when a heterologous polypeptide or amino acid is inserted into the polypeptide. The scaffold provides for maintenance of a structural and/or functional feature of the polypeptide after the heterologous polypeptide has been inserted. In one embodiment, a scaffold

comprises an antibody variable domain, and maintains a stable structure when a heterologous CDR or amino acids are inserted into the scaffold.

[75] The term "library size" as used herein refers to the number of phages that comprise the library. Several methods can be used to estimate the library size, being the most common counting the number of antibiotic-resistant colonies after plating a dilution of the library. Following the electroporation, transformed bacteria are incubated in rich medium for no more than 55 min before plating dilutions on agar supplemented with the appropriate antibiotic. This procedure ensures that the original transformants are counted before the bacteria start to divide actively.

[76] The term "functional space" or "effective size" of a library as used herein refers to the number of unique antibody sequences in a library that produce functional antibody fragments fused to a phage particle. For instance, nucleotide sequences with stop codons like UAA ("ochre") or UGA ("opal") lead to truncated sequences and do not produce functional antibody fragments fused to a virion particle. Insertions or deletions of one or two nucleotides change the reading frame of the gene sequence leading to stretches of amino acids that may impair folding or lead to non-functional clones.

[77] The functional space of a library can correspond with the library size if all the sequences are functional or is a fraction of the library size. The functional space can also be called 'shape space' or 'sequence space'.

[78] The term "stability" as used herein refers to the ability of a molecule to maintain a folded state such that it retains at least one of its normal functional activities, for example, binding to an antigen or to a molecule like Protein A. The stability of the molecule can be determined using standard methods. For example, the stability of a molecule can be determined by measuring the thermal melt ("Tm") temperature. The Tm is the temperature in degrees Celsius at which ½ of the molecules become unfolded. Typically, the higher the Tm, the more stable the molecule.

[79] As used herein, "natural" or "naturally occurring" polypeptides or polynucleotides refers to a polypeptide or a polynucleotide having a sequence of a polypeptide or a polynucleotide identified from a no synthetic source. For example, when the polypeptide is an antibody or an antibody fragment, the no synthetic source can be a differentiated antigen-specific B cell obtained *ex vivo*, or its corresponding hybridoma cell line, or from the serum of an animal. Such antibodies can include antibodies generated in any type of immune response, either natural or otherwise induced. Natural antibodies include the amino acid sequences and the nucleotide sequences that constitute or encode these antibodies, for example, as identified in the Kabat database. As used herein, natural antibodies are different than "synthetic antibodies", synthetic antibodies referring to antibody sequences that have been changed, for example, by the replacement, deletion, or addition, of an amino acid, or more than one amino acid, at a certain position

with a different amino acid, the different amino acid providing an antibody sequence different from the source antibody sequence.

[80] A "plurality" or "population" of a substance, such as a polypeptide or polynucleotide of the invention, as used herein, generally refers to a collection of two or more types or kinds of the substance. There are two or more types or kinds of a substance if two or more of the substances differ from each other with respect to a particular characteristic, such as the variant amino acid found at a particular amino acid position. In a non-limiting example, there is a plurality or population of polynucleotides of the invention if there are two or more polynucleotides of the invention that are substantially the same, preferably identical, in sequence except for one or more variant amino acids at particular CDR amino acid positions.

[81] The term "developability" as herein used, refers to a set of desirable antibody properties which, as a whole, facilitates clinical development and manufacturing of a therapeutic antibody. Properties that have been associated with developability include, but are not limited to, good expression (>30 mg/L in transient CHO cell expression), thermal stability (>70°C), low or no aggregation (< 1% high molecular weight aggregates; HMWA), and solubility of > 40 mg/ml for IV indications and > 100 mg/ml for subcutaneous indications.

[82] These properties are interrelated. For instance, chemical instabilities such as oxidation or clipping sites result in sample heterogeneity and eventually can impact the physical stability or lead to low solubility or aggregation. Poor physical stability can expose side-chains prone to oxidation or degradation, eventually leading to aggregation when these residues are degraded. Nevertheless, each of these parameters can be measured independently. For instance, expression can be assessed by measuring the amount of antibody in the culture media by ELISA, Octet™ or BIAcore™. Thermal stability can be measured by Thermal shift analysis or DSC. Aggregation can be measured by SEC-HPLC. Solubility can be measured by concentrating the antibody at diverse concentrations.

[83] Molecules that do not meet the success criteria above described are commonly deprioritized in the therapeutic antibody development process. Alternatively, if antibody leads were identified with promising biological activity but low performance during the developability assessment, they are targeted for developability enhancement by the methods described in the art during the optimization phase.

[84] "Phage display" is a technique by which variant polypeptides are displayed as fusion proteins to at least a portion of a coat protein on the surface of phage, e.g., filamentous phage, particles. A utility of phage display lies in the fact that large libraries of randomized protein variants can be rapidly and efficiently sorted for those sequences that bind to a target molecule with high affinity. Display of peptide and protein libraries on phage has been used for screening millions of polypeptides for ones with specific binding properties. Polyvalent phage display methods have been used for displaying small random

peptides and small proteins through fusions to either gene III, VIII or IX of filamentous bacteriophage, see references cited therein.

[85] In monovalent phage display, a protein or peptide library is fused to a gene III or a portion thereof, and expressed at low levels in the presence of wild type gene III protein so that phage particles display one copy or none of the fusion protein. Avidity effects are reduced relative to polyvalent phage so that sorting is on the basis of intrinsic ligand affinity, and phagemid vectors are used, which simplify DNA manipulations.

[86] A "phagemid" is a plasmid vector having a bacterial origin of replication, e.g., ColE1, and a copy of an intergenic region of a bacteriophage. The phagemid may be used on any known bacteriophage, including filamentous bacteriophage and lambdoid bacteriophage. The plasmid will also generally contain a selectable marker for antibiotic resistance. Segments of DNA cloned into these vectors can be propagated as plasmids. When cells harboring these vectors are provided with all genes necessary for the production of phage particles, the mode of replication of the plasmid changes to rolling circle replication to generate copies of one strand of the plasmid DNA and package phage particles. The phagemid may form infectious or non-infectious phage particles. This term includes phagemids, which contain a phage coat protein gene or fragment thereof linked to a heterologous polypeptide gene as a gene fusion such that the heterologous polypeptide is displayed on the surface of the phage particle.

[87] As used herein, "amplify" refers to the process of enzymatically increasing the amount of a specific nucleotide sequence. This amplification is not limited to but is generally accomplished by PCR. As used herein, "denaturation" refers to the separation of two complementary nucleotide strands from an annealed state. Denaturation can be induced factors such as, for example, ionic strength of the buffer, temperature, or chemicals that disrupt base pairing interactions.

[88] The terms "amplification cycle" and "PCR cycle" are used interchangeably herein and as used herein refers to the denaturing of a double-stranded polynucleotide sequence followed by annealing of a primer sequence to its complementary sequence and extension of the primer sequence.

[89] The terms "polymerase" and "nucleic acid polymerase" are used interchangeably and as used herein refer to any polypeptide that catalyzes the synthesis or sequencing of a polynucleotide using an existing polynucleotide as a template.

[90] As used herein, "DNA polymerase" refers to a nucleic acid polymerase that catalyzes the synthesis or sequencing of DNA using an existing polynucleotide as a template.

[91] All publications and patents mentioned herein are incorporated herein by reference for all purposes, including the purpose of describing and disclosing, for example, the constructs and methodologies that are described in the publications, which might be used in connection with the presently described invention. The publications discussed herein are provided solely for their disclosure

prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention or for any other reason.

[92] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices, and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

DETAILED DESCRIPTION OF THE INVENTION

[93] This invention includes methods to generate highly functional antibody libraries. In one embodiment, the methods comprise the steps of: (i) generation of a primary library (PL) of antibody fragments; (ii) a process to select for a well-folded antibody fragment library (FL) from the PL; and (iii) combining the antibody fragments selected in step 2 with diverse CDR-H3 fragments to generate a highly functional secondary antibody library (SL). A detailed description of modalities these steps and exemplary proofs of each step follow.

Primary Library (PL) Generation

[94] To demonstrate the methods in the present invention, two primary libraries were built with human antibody germline genes. Both libraries had a universal V_H scaffold and two V_L scaffolds. The V_L scaffolds provided two alternative antigen-binding site topographies. One of the V_L scaffolds had a short L1 loop (PL1), whereas, the other had a long L1 loop (PL2) (**Figure 1**). By changing the length of L1 from a short to a long loop, antibodies alter the preference to bind protein or peptide targets (Vargas-Madrazo et al., 1995). Therefore, by using the proper V_L scaffold, the repertoires exemplified herein can be used for selection of antibodies against protein or peptide targets. In one embodiment, libraries of the present invention built on these scaffolds were used in combination to generate antibodies that bind diverse types of epitopes on a given target.

[95] As shown in Example 1, the V_L scaffold with a short L1 loop was built by assembling the human IGKV3-20*01 germline gene with the human IGJV4*01 germline gene (SEQ ID NO: 1). The V_L scaffold with the long L1 was built by assembling the human IGKV4-01*01 germline gene combined with the same human IGJV4*01 germline gene (SEQ ID NO: 2). IGKV3-20*01 and IGJV4*01 belong to different IGKV families and share only 68% identities. Thus, the scaffolds built with these two genes

represent two distantly related human germline genes and thus offer distinct exemplary proofs of this invention.

[96] As counterpart of the V_L scaffolds, a single V_H scaffold, built with the human IGHV3-23*01 and the human IGJH3*01 germline genes (SEQ ID NO: 3) was used. In one aspect of this invention, Protein A binds the framework region 3 (FR-3) of the V_H domain encoded by germline gene IGHV3-23*01. The FR-3 is formed by discontinuous amino acid stretches distant in the primary sequence but brought together by folding. Therefore, Protein A has been used in prior art for its ability to bind only to well-folded V_H domains (Jespers et al., 2004).

[97] Other natural ligands that bind folded variable regions of antibodies outside the antigen-binding site have been described in the art. For example, Protein-L binds the V_L domain of antibodies encoded by the human IGKV-1, IGKV-2 and IGKV-4 gene families (Nilson et al., 1992). More specifically, the IGKV4-01*01 germline gene, which belongs to the human IGKV4 family, binds Protein-L. The use of Protein-L as a ligand to select well-folded antibodies alone and/or in conjunction with Protein A is a clear extension of this invention. For instance, human V_H domains of antibodies encoded by the gene families IGHV-1, IGHV-2, IGHV-4, IGHV-5, IGHV-6 and IGHV-7, which do not bind Protein A, can be paired with libraries of V_L domains encoded by scaffolds built with members of the human IGKV-1, IGKV-2 and IGKV-4 gene families. These libraries and antibody pairs can then be submitted to diverse destabilizing conditions to select for well-folded antibodies with Protein-L.

[98] Yet in another aspect of the invention, the germline genes used to build the V_H and V_L scaffolds of this invention have frequently been observed in human antibodies elicited against a vast array of diverse antigens (Nilson et al., 1992). These genes have also been used as foundation to build numerous scFv and Fab libraries for antibody discovery in the art (US patent 9,062,305) (Shi et al., 2010), as well as in humanization of therapeutic antibodies (US patent 8,777,044). Therefore, it is expected that antibodies discovered from libraries built with these scaffolds will perform well both *in vitro* and *in vivo* settings and will be amenable to further therapeutic development.

[99] In other aspects of the invention, the Tm of the $_H$Vand V_L scaffolds in Fab format has been measured by Differential Scanning Calorimetry (DSC) yielding similar values of 75°C (Teplyakov et al., 2016). This Tm is almost 10°C above the Tm of the C_H2 domain, which is estimated at 68°C (Gilliland et al., 2012). The C_H2 is the least stable domain of the human IgG1 molecule and hence the first domain that unfolds. Therefore, the antibodies isolated from the libraries herein described are expected to be highly stable.

[100] In yet in another related aspect of the invention, antibodies encoded by the exemplary V_H and V_L scaffolds herein disclosed have been solved by x-ray crystallography (Teplyakov et al., 2016) in association the CDR-H3 loop of a known therapeutic antibody, CNTO 888 (Obmolova et al., 2012). This

knowledge facilitated the design of diversity in all the CDRs that form the antigen-biding site except in the CDR-H3.

[101] The diversity of the V_H scaffold was focused on the CDR-H1 and CDR-H2 and was designed at positions and amino acids commonly found in contact with protein and peptide antigens, positions accessible to antigens in the antigen-binding site and/or in contact with the CDR-H3. As exemplified in Example 2, to determine the positions in contact with protein and peptide targets, all the antibody structures compiled at the PDB and curated by the IMGT as of March 17 th, 2017 were analyzed. The initial dataset consisted of 2,645 antibody structures from diverse species and specificities. From this initial dataset, only anti-protein antibodies with different names and those solved at $\leq$ 3Å resolution were considered for identifying positions to diversify. In addition, only the antibody structures with the same length at CDR-H1 and CDR-H2 were studied. Although this filter is certainly stringent, it removes for the most part antigen-binding site structures with canonical structures classes (Chothia and Lesk, 1987) other than that of the V_H scaffold. Having the same canonical structures guarantees the same relative position of the amino acids at the antigen-binding site to contact antigens. As additional precaution to design the diversification of the V_H scaffold, the V_H sequences with the same length were compared by a clustering algorithm to remove those showing 100% of identity and thus avoided counting the same sequence twice. The final dataset contained 117 antigen:antibody complexes which are listed in **Figure 2**.

[102] To identify the positions in contact with antigens, the contact tables of the 117 structures listed in Figure 2 were downloaded from the IMGT and the contact residues aligned. The residues in contact with antigens were then mapped onto the structure of the V_H scaffold (PDB ID: 5DIL) to determine their ASAs and structural environment. A total of 10 positions were targeted for diversification, four in the CDR-H1 and six in the CDR-H2. These positions have >70% ASA and thus should tolerate diverse amino acid side chains without disruption of the V_H scaffold. On other hand, all targeted positions but 30 at the CDR-H1 and 55 at the CDR-H2 were found in > 30% contacts, thus maximizing the probability of interacting with diverse antigens. Positions 30 and 55 are in the periphery of the antigen-binding site and are highly exposed to the solvent (> 65% ASA). Hence, although these positions are in contact antigens in less than 50% of the complexes, they were considered for diversification to supplement diversity for binding targets of bigger size than average proteins.

[103] The diversification regimes at CDR-H1 and CDR-H2 were designed using three sources of information: the dataset of curated antibody structures listed in **Figure 2**, the V_H sequences available at NCBI and curated by Abysis, and the germline genes of the human IGHV3 family compiled at the IMGT. The number of V_H sequences available at Abysis amounts 76,000 V_H sequences, thus complementing the information obtained from the two other sources. The comparison with the frequency of amino acids

encoded in an alignment of the human IGHV3 germline family ensured that the amino acids used for diversifying the scaffolds mimicked the human germline diversity.

[104] The amino acids targeted for diversification were examined in the structure of the primary library scaffolds (**Figure 1**) to avoid conflicts and/or inconsistencies between the nature of amino acids used for diversification and the environment of such amino acids in the context of the scaffolds. For instance, residues buried in the protein had to be hydrophobic. Residues in β-turns or maintaining the canonical structures such as position 54 at the CDR-H2 had to be consistent with the propensity of a residue in such conformation. Aromatic residues had to be close to other aromatic residues or residues with relatively long aliphatic side-chains. The final diversification regime at CDR-H1 and CDR-H2 of the V_H scaffold is summarized in **Figure 3**. The estimated number of amino acids per position is listed in the last column of the figure and yielded a diversity of 5.4 x 10^9 unique amino acid V_H sequences. Some positions are relatively conserved such as position 30 at the CDR-H1 with 2 amino acids allowed, whereas other positions are heavily diversified with up to 9 residues, e.g., position 50 of CDR-H2.

[105] To identify positions for diversification in the V_L scaffolds, the same procedure as in the V_H scaffold was followed. First, positions in contact with the antigen were identified in the dataset of 117 structures listed in **Figure 2**. Next, positions in contact with the antigen were mapped onto the structure the V_L scaffolds (**Figure 1**) to evaluate their solvent exposure, structural context and interaction with the V_H Scaffold.

[106] As in the V_H scaffold, the diversification regime of the V_L scaffolds was first determined by the frequency of amino acids occurring in the alignment of the 117 structures (**Figure 2**). This information was supplemented with the analysis of all the V_L sequences compiled by Abysis, total 20,764 sequences. This information was then compared with the amino acids and frequencies of the human IGKV germline gens compiled at the IMGT. Fewer V_L than V_H sequences at Abysis reflects the fact that ~ 60% of the human antibody repertoire is k-type, and only k-type sequences are considered in design. Finally, the diversification regime of the V_L scaffolds was evaluated in the structures of V_H and V_L scaffolds (**Figure 1**). The final diversification regime of the V_L scaffolds is summarized in **Figure 4** and **Figure 5**. The estimated number of amino acids per position is listed in the last column of each of the two figures and yielded a total of 1.1 x 10^6 and 1.3 x 10^6 unique amino acid V_L sequences for PL1 and PL2, respectively.

[107] In other aspect of the invention, it was reasoned that by using only one CDR-H3 sequence to generate the PLs, for instance the CDR-H3 of CNTO 888, the diversity of amino acids in contact with, or nearby, said CDR-H3 may be constrained to a few and specific residues to accommodate said CDR-H3 under specific selection conditions. Therefore, a set of CDR-H3 sequences called "neutral H3Js" was designed by starting from the repertoire of human IGDH genes and combining them with the human IGJH segments in germline configuration. Since antibodies from natural primary repertoires are mostly

sequences in germline configuration, it was hypothesized that the neutral H3Js provide enough diversity to avoid biases in amino acids in contact with, or nearby, the neutral H3J sequences while contributing to select for developable antibodies. As exemplified in Example 3, the set of neutral H3Js fragments comprises ninety (90) fragments (SEQ ID Nos: 4-93).

[108] The designs of the diversified V_H and V_L scaffolds of this invention with the neutral H3Js were assembled in the configuration depicted in **Figure 6**. The V_L-linker-V_H scFv configuration places the H3J fragments on the C-terminal. V_L is physically linked to V_H by a linker of the 19 amino acids GGGGSGGGGSGGGSGGGGS (GS19). Two BglI/SfiI sites on each side allows for in-frame cloning in the acceptor vector between the pelB leader peptide for periplasmic expression and the tags for detection. Other configurations including V_H-linker-V_L and other linker sequences are obvious extensions of this invention.

[109] In one embodiment of the invention, the NcoI restriction site and the KpnI restriction site are common to both repertoires, allowing for easy replacement of V_L. In other embodiments, the JK anchor sequence in V_L and the human VH Consensus Motif (CM) in V_H (Example 8) are used for exchange and cloning of natural H3J fragments in the third step of the invention or for exchanging the light chains. These two short sequence stretches are common to PL1 and PL2.

[110] The resultant scFv designs were synthesized using trimer phosphoramidite technology. Nevertheless, other various methods described in the art can be used to realize the designs, including mix of oligonucleotides and restriction enzymes. For instance, seamless DNA assembly using type IIs enzymes, e.g. Golden Gate assembly (New England Biolabs) can be used to concatenate V_H and V_L diversified scaffolds and H3Js fragments in a single piece of DNA.

[111] The synthetic fragments of this invention were cloned as fusions to pIII by digestion with BglI restriction enzyme and ligated into pADL™-23c phagemid vector (Antibody Design Labs, San Diego, cat# PD0111) as described in Example 4. The pADL™-23c phagemid vector is a classical type 3+3 phage display vector with a cloning site for display on the N-terminal side of the full-length gene III protein. Secretion in the periplasm of the fusion protein is driven by the PelB leader peptide. Display of the scFvs on the phage is obtained with the help of an amber-suppressive bacterial strain and M13KO7 helper phage. Growth on non-suppressive strains results in the expression of free scFvs in the periplasm space, which can be subsequently purified by immobilized metal affinity chromatography (IMAC) with the help of the His tag and detected with the Myc tag as described in prior art.

[112] The initial diversity of PL1 and PL2 was 1.7×10^9 and 2.3×10^9 cfu, respectively. To prove that PL1 and PL2 displayed as fusion proteins on the phage surface bound Protein A and hence can be submitted to the step 2 of the invention, the libraries were assayed on Protein A by a direct phage ELISA

(Example 5). **Figure 7** shows that both primary libraries bound to Protein A, with PL2 showing a higher EC_{50} than PL1.

[113] Although it was considered that these examples were the best modality in the field of this invention, there is no limitation to the origin of the PL, besides the understanding noted above that some constructions, use of specific scaffolds or natural sources may be more suitable for developable antibody generation. For instance, a PL can be obtained from natural sources such as PBMCs and lymphoid organs (lymph nodes, spleen), either from human or animals that have not been immunized against any specific antigen, hyper-immunized animals or individuals suffering or not from a debilitating condition or circumstance of immunological interest such as a recent vaccination or an infectious episode.

Selecting for Highly Stable Fragments and FLs

[114] In this step, the PLs are submitted to diverse conditions to select for highly stable and developable antibody variants. In the best modality of this invention, heat was used to eliminate unstable variants from the pool of primary antibodies. It has been shown that transient heat treatment of antibodies displayed on phage can lead to denaturation and aggregation of the less thermostable antibody variants in a library (Jespers et al., 2004). Therefore, following heat–induced denaturation and aggregation of variants with lower stability profiles, rescue by Protein A through direct binding yielded filtered libraries of improved functionality such as in-frame and thermostable clones. A well-known indirect effect to those skilled in the art of the process of selection and re-amplification of phage libraries as operated in the above filtration is the drastic decrease of diversity that can lead to the rapid collapse of libraries (Matochko et al., 2014). The main purpose of this invention is to provide a method to improve the diversity, hence functionality, of antibody library filtrated for improved developability without impacting the quality of the antibody variants collected along the filtration process.

[115] Conditions for selecting developable and stable antibody variants from PLs include, but are not limited to heat. Conditions capable to induce unfolding, or combination of thereof, are obvious extensions of this invention. For example, although without clear literature examples, but obvious from the skilled in the art, high or low pH, high salt concentrations or any chaotropic conditions such as urea may be used to interrogate antibody stability. In another application, partial unfolding of the antibody structure may lead to exposure of a stretch of the polypeptide backbone and protease sensitivity. This approach has been used to engineer antibody with high-protease resistance capable to survive the stress conditions found in the gastrointestinal track and is a clear extension of this invention (Hussack et al., 2011).

[116] Another fundamental aspect of antibody reactivity is the non-specificity or multi-specificity. Non-specific, multi-reactive B-cell clones are eliminated during the maturation of the B-cell repertoire. Use of non-specific absorption and hydrophobic interaction may replicate such a process *in vitro* and are an obvious extension of this invention. In this case, rather than applying a positive selection of well-folded antibodies, e.g. with Protein–A, unwanted clones are passively eliminated from the pool. A cumbersome alternative to assay for non-specificity is to test each antibody lead along the pipeline against panels of antigens and proteins. Tools to predict antibody non-specificity applicable to entire libraries such as binding to chaperones, e.g., HSP70 or HPS90 (Kelly et al., 2017) are a clear extension of this invention.

[117] In an embodiment of this invention, well-folded antibody variants are rescued with a ligand that selectively binds said well-folded antibodies. As shown in Example 5, one such a binder is Protein A. Herein is exemplified this critical step with incubation of PL1 and PL2 in a range of temperatures followed by the rescue of the folded variants by Protein A. To set up the optimal conditions for selection of developable antibody fragments, Example 6.1 shows the incubation for 10 min of the 3-20/3-23 and 4-01/3-23 scFvs prior to diversification. The scFvs displayed as fusions to the minor phage coat protein pIII on the phage surface were submitted for a range of temperatures, starting at 40°C and increasing the temperature up to 80°C in steps of 5°C. The unfolding process was monitored with a direct Protein A ELISA (**Figure 8**). Since the M13 phage is stable at 80°C, a change in the ELISA signal in this range of temperatures is a direct consequence of the scFv unfolding.

[118] The unfolding process, which leads to aggregation of the scFv-phage and hence a drop in the ELISA signal depends on two parameters (Jespers et al., 2004): (i) thermal stability of the scFvs, and (ii) number of copies of the scFvs displayed on the phage surface, which varies from one to five copies. Once the scFv starts to unfold, a cooperative aggregation process takes place. ScFv-phages with a higher number of copies aggregate first, serving of aggregation seed for phages with a fewer number of scFv copies.

[119] Example 6.1 indicated that the 3-20/3-23 scFv started to unfold at 65°C, whereas, the 4-01/3-23 scFv started the unfolding process at 55°C. The Tm of the 3-20/3-23 scFv is 75°C, at which 50% was unfolded. The Tm of 4-01/3-23 scFv is 65°C. The unfolding kinetic at 60°C and 72°C of the 3-20/3-23 scFv and 4-01/3-23 scFv demonstrated that the 3-20/3-23 scFv remained folded at 60°C for up to one hour (**Figure 9**). At 72°C it unfolded slowly with approximately a 20% drop in the ELISA signal during the first 30 min. Afterwards, a quick drop of the ELISA signal was seen. The 4-01/3-23 scFv unfolded very slowly at 60°C. At 72°C it showed a quicker unfolding process than 3-20/3-23 scFv, with a drop in the signal of approximately 30% in the first 10 min. This process was accelerated afterwards, probably due to a massive aggregation of phage with a fewer number of 4-01/3-23 scFv copies.

[120] Therefore, this example demonstrated that: (i) the 3-20/3-23 and 4-01/3-23 scFvs used as scaffolds to generate the exemplary PLs of this invention are stable at 60°C for at least one hour; (ii) said 3-20/3-23 and 4-01/3-23 scFvs can be incubated at 72°C for 10 min without significant unfolding and aggregation, i.e., < 30%; (iii) the 3-20/3-23 and 4-01/3-23 scFvs have specific Tm's and dynamic of unfolding, which are intrinsic properties of the VL scaffold sequence and structure and can be used to tailor the selection conditions to generate a variety of developable antibodies by changing the harsh denaturing conditions.

[121] Accordingly, and as shown in Example 6.2, the exemplary PLs of this invention produced unique scFvs with a significant number, on average ~52.3% (Table 3), of well-folded antibodies after incubation at 72°C for 10 min. This number was down from on average ~64% in both PLs prior to the heat treatment, representing a combined number of more than 2 billion folded but unstable antibodies for PLs of 1 x 10^{10} clones or more each.

[122] Considering that the Tm of the C_H2 domain of the human IgG1 is 68°C, as an example of selection for highly stable and developable antibodies in this invention, it was decided to submit PL1 and PL2 for incubation at 70°C for 10 min and rescue well-folded antibody variants with Protein A as shown in Example 7. The resulting libraries, FL1 and FL2, had both the hallmark of a selection by Protein A. All single clones, but one in the respective sampling, were in-frame, giving a percentage of in-frame clones of 94.7% and 97.3% in FL1 and FL2, respectively. Longer or shorter incubation times at higher or lower temperatures are obvious extensions of the method herein disclosed.

[123] The use of Protein A to improve the functionality of antibody libraries is a well-described procedure, but is known to result in a significant loss of diversity (Jespers et al., 2004; Famm et al., 2008; Rouet et al., 2014). For instace, antibody domains libraries displayed on filamentous phage have been submitted to 80°C in pH 7.4 by Jesper et al., and folded domains have been selected by binding to Protein A. Using the same phage library repertoire, Famm et al. have extended the method to the selection of domains resistant to acid aggregation. In both cases, as expected, the number of stable clones have been reduced dramatically.

Construction of Secondary Libraries (SLs)

[124] The present invention also includes methods to restore diversity to an antibody library that has lost some of its diversity due to enrichment and/or selection. By combining the highly stable variants selected in the previous step with a collection of natural CDR-H3 fragments the diversity of the libraries can be restored. Two SL libraries, SL1 and SL2, were generated starting from plasmid DNA of FL1 and FL2 as substrate for amplification of well-folded scaffolds by PCR as exemplified in Example 8. The

PCR-generated fragments included the pelB leader peptide-encoding nucleotide sequence and a stretch of nucleotides immediately before the CM (**Figure 16**). The sequence of the CM allowed the amplification of more than 95% of all antibody sequences found in circulating PBMCs as described in Example 8. During the amplification, a BsaI restriction site was added immediately after the end of the V_H fragment. Symmetrically, a repertoire of natural H3Js fragments was obtained from 200 healthy donors as described in Example 8.1. It was generated with a pair of primers matching the sequence of the CM and the pADL™-23c phagemid vector sequence downstream to the second SfiI site. In doing so, a BsaI site was added 5' to the CM. Simultaneous digestion by BsaI of the two fragments and ligation led to the joining of the FLs with the natural H3J fragments (**Figure 17**). Subsequent amplification of complete scFvs by nested primers led to the successful cloning of very large SL1 and SL2 libraries (Example 8.2). The initial diversity was 1.4×10^{10} and 1.1×10^{10} cfu for SL1 and SL2, respectively. The use of the digestion/ligation joining reaction enabled the subsequent amplification starting from large amounts of well-assembled full-length scFvs. Other methods of molecular biology well-known to those skilled in the art, such as PCR by overlapping extension, are obvious extensions of the method to build the SL libraries.

[125] The quality of the SLs was assessed by Sanger sequencing as described in Example 8.3. It was found that 35 out of 41 clones (85.4%) were in-frame for both SL1 and SL2. These values were only on average 10% lower than those of the FLs, indicating that our method of the SLs construction retained the vast majority of the in-frame clones from the FLs. In addition, all the sequence studied happened to have unique natural CDR-H3 sequences. Therefore, the successful addition of natural H3J diversity into the FLs was performed to produce the SLs while retaining the in-frame character resulting from the filtration.

[126] When single clones from SL1 and SL2 were expressed as scFv on the phage surface, and after incubation at 72°C for 10 minutes, over 80% of them bound Protein A (**Figure 20** and **Figure 21**). The comparison of the percentage of stable clones between PLs and SLs showed an increase in absolute value of 15% to 20%, equivalent ~ 1.5×10^9 to 2×10^9 additional variants in a library of 10^{10} variants (**Figure 22**). In relative values, the percentage of stable clones grew by an average of ~33% in both SLs comparatively to the PLs using the library construction method of this invention. Since PLs are representative examples of typical synthetic libraries in the art, the newer created libraries of this invention, or SLs, expanded the functional space of the previous synthetic libraries. This improvement increased significantly the probability of isolating specific and higher affinity antibodies from the SLs.

[127] Most of the natural CDR-H3 loops are in-frame. Therefore, it was expected that most of the clones from the SLs resulted in in-frame scFv sequences. It was unexpected, that ~70% of the SL clones survived the incubation for 10 min at 70 °C. Natural CDR-H3 fragments have not been exposed to such harsh conditions in physiological conditions, and hence have not evolved to be stable at such a high temperature. Moreover, it was not predictable from prior art that the natural CDR-H3 would interact well

with the stable scaffolds from the FLs. The natural CDR-H3 fragments of this invention were isolated from PBMCs of a pool of 200 donors and amplified with primers to generate around 95% of all the CDR-H3 in those individuals. The CDR-H3 is highly variable, with length variation between 3 and more than 20 amino acids, recombined with over 40 IGHV and 6 IGHJ and paired with ~ 40 IGKV genes. This enormous diversity when cloned in the highly stable exemplified scaffolds of the invention, only impacted ~10% of the overall library diversity while improving stability. Therefore, it is likely that an interplay between the selection among the human germline genes of highly stable scaffolds to build the libraries, with the designed diversification regime based on germline genes, compounded with removal of liability developabilities, expressed in conjunction with the set of neutral H3Js designed to be flexible and finally, the selection of PLs, produced to a collection of highly stable antibody variants suited to accommodate such a diverse natural collection of CDR-H3 fragments.

[128] As shown in Example 9, analysis of the libraries generated during the construction by next generation sequencing (NGS) confirmed the quality of PLs in accord with the intended design. Analysis of the filtration process showed as expected an increase in productive sequences and a decrease of the diversity after the filtration process. Distribution of the neutral D elements and JH fragments showed limited bias, with the exception of a trend toward the elimination of the most hydrophobic fragments, underlining the need for a varied local environment around CDR-H3 during the filtration process to minimize biases in the diversity as provided by the set of neutral H3J fragments. NGS analysis of the SL confirmed the presence of very high levels of diversity at both CDR-H3 level, with around 60% unique sequences, and full antibody sequence level with above 99% diversity at the depth of around one million sequences in each SL (**Table 12**). Analysis of the most prevalent CDR-H3 indicated a limited copy number for each clone, to the contrary of what one would expect from re-amplified libraries.

[129] As shown in Example 10, standard selection techniques applied with two known antigen models, TNF and HSA , were performed. In both examples specific antibodies were isolated, demonstrating the potential of the SLs to produce specific antibodies

EXPERIMENTAL EXAMPLES

[130] The following examples are offered to illustrate, but not to limit the claimed invention. It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All

publications, patents, and patent applications cited herein are hereby incorporated by reference herein in their entirety for all purposes.

EXAMPLE 1: DESIGN OF TWO HUMAN SCAFFOLDS FOR RECOGNITION OF DIVERSE EPITOPES

[131] To maximize the functionality of the PLs, antibody scaffolds highly utilized in human with high expression profile and unique recognition properties were selected from the human germline repertoire. Two PLs were built with a V_L scaffold that has a short L1 loop (SEQ ID NO: 1) and a V_L scaffold that has a long L1 loop (SEQ ID NO: 2). As counterpart of the V_L scaffolds, a single V_H scaffold (SEQ ID NO: 3) was used. By changing the length of L1 from a short to a long loop, antibodies alter the preference to bind protein or peptide targets, respectively.

EXAMPLE 2: DESIGN OF DIVERSIFICATION REGIMES FOR TWO HUMAN SCAFFOLDS

[132] The diversity of the V_H scaffold was focused on the CDR-H1 and CDR-H2 and was designed at positions and amino acids commonly found in contact with protein and peptide antigens, positions accessible to antigens in the binding site and/or in contact with CDR-H3. To determine the positions in contact with protein and peptide targets, 117 antigen:antibody complexes listed in **Figure 2** were used. The positions in contact with antigens were identified by downloading the contact tables of the 117 structures from the IMGT and aligning the contact residues. Then, the residues in contact with antigens were then mapped onto the structures of the V_H scaffold paired with the two V_L scaffolds (PDB IDs: 1ILC and 1ILD) to determine their ASAs and structural environment.

[133] The diversification regime at CDR-H1 and CDR-H2 was designed using three sources of information: The dataset of curated antibody structures listed in **Figure 2**, the V_H sequences available at NCBI and curated by Abysis, and the germline genes of the human IGHV3 family compiled at the IMGT. The number of V_H sequences available at Abysis amounted for 76,000 V_H sequences, thus complementing the information obtained from the two other sources. The final diversification regime at CDR-H1 and CDR-H2 of the V_H scaffold is summarized in **Figure 3**. The estimated number of amino acids per position is listed in the last column of **Figure 3** and yields a diversity of 5.4×10^6 unique amino acid V_H sequences.

[134] To identify positions for diversification in the V_L scaffolds, we followed the same procedure as for the V_H scaffold. First, positions in contact with the antigen were identified in the dataset of 117 structures listed in **Figure 2**. Next, positions in contact with the antigen were mapped onto the structure of the V_L scaffolds (**Figure 1**) to evaluate their solvent exposure, structural context and interaction with the V_H scaffold. The diversification regime was first determined by the frequency of amino acids occurring in the alignment of the 117 structures (**Figure 2**). This information was supplemented with the analysis of all the V_L sequences compiled by Abysis, totaling 20,764 sequences. This information was then compared with the amino acids and frequencies of the human IGKV germline genes compiled at the IMGT. The final diversification regime of the V_L scaffolds is summarized in **Figure 4** and **Figure 5**. The estimated number of amino acids per position is listed in the last column of **Figure 4** and **Figure 5** and yields a total of 1.1×10^6 and 1.3×10^6 unique amino acid V_L sequences for the 3-20 and 4-01 repertoires, respectively.

EXAMPLE 3: NEUTRAL CDR-H3 DESIGN

[135] To design the neutral H3Js, the IGDH germline genes compiled at the IMGT were translated in the three reading frames into amino acid sequences. Only productive sequences, i.e. without stop codons, were considered. Those IGDH genes with developability liabilities including methionine (M) and tryptophan (W) residues, as well as those encoding three or more hydrophobic residues were removed from the set of neutral H3Js. After this *in silico* selection process, 18 sequences from the IGDH germline genes were combined with 5 IGJH regions to produce ninety (90) fragments (Table I) (SEQ ID Nos: 4-93). The length of the neutral H3Js varies between 18 and 27 amino acids.

TABLE I

"Neutral" D Elements	J Regions
GYSGYDY	AEYFQHWGQGTLVTVSS
GYSYGY	DAFDVWGQGTMVTVSS
TTVT	YFDYWGQGTLVTVSS
YSGSYY	NWFDSWGQGTLVTVSS
DYGDY	YGMDVWGQGTTVTVSS
DYSNY	
SIAAR	

VQLER

YYDILTGYYN

YYYDSSGYYY

YYYGSGSYYN

EYSSSS

GITGT

GIVGAT

GTTGT

GYSSGY

LTG

VDIVATI

EXAMPLE 4: SYNTHESIS AND CLONING OF PRIMARY LIBRARIES

[136] In this example, the two scFv libraries with the diversity reported in **Figures 3-5** and the configuration depicted in **Figure 6** were submitted for chemical synthesis using trimer phosphoramidite mixtures at the position of degeneracy. The set of neutral H3J fragments (SEQ IDs: 0004-0093) was incorporated on the 3' side of the synthetic fragments. The quality of the synthetic fragments was assessed by Sanger sequencing of 96 clones taken randomly from the final synthetic fragments. The results are summarized below (Table 2):

TABLE 2

Repertoire	**3-20/3-23**	**4-01/3-23**
Sample Size	96 clones	96 clones
In-frame Clones	62%	73%
Design Compliant Clones	50%	60%

[137] The 3-20/3-23 synthetic fragments exhibited 50% of sequences in-frame matching the design, whereas the 4-01/3-23 fragments had 60%. Around 10% of the fragments had either an in-frame insertion

or a deletion. Although these sequences did not comply with the intended design, they could positively contribute to the diversity of the libraries.

[138] The expected and observed frequency of the IGHJ usage was evaluated in the in-frame sequences in both libraries and showed good correspondence between the designed and observed frequency. Comparison of the expected vs observed frequency of the amino acids per diversified position was also conducted. Good agreement was found, with differences falling within the expected variations of trimer-based oligonucleotide synthesis.

[139] The two libraries were cloned starting from one microgram of each synthetic fragment 3-20/3-23 or 4-01/3-23, to generate PL1 or PL2, respectively. Digested fragments with BglI restriction enzyme overnight at 37°C were ligated into pADL™-23c phagemid vector (Antibody Design Lab, San Diego). The ligation reactions were electroporated into electro-competent TG1 cells (Lucigen) and transformants were rescued on 2xYT medium supplemented with ampicillin at 37°C in the presence of glucose 1% w/v. The initial diversity was 1.7×10^9 and 2.3×10^9 primary transformants for PL1 and PL2, respectively.

[140] Cells were harvested after overnight incubation, resuspended in fresh 2xYT medium supplemented with ampicillin and subsequently superinfected with M13KO7 helper phage. No more than 55 min after transduction, kanamycin 50 μg/ml was added, the temperature was lowered to 30°C and the incubation was prolonged overnight. The morning after, virions were purified by PEG precipitation following standard protocols to those skilled in the art and references herein described.

[141] Sanger sequencing of PL1 and PL2 clones selected at random from the primary transformants showed no background vectors in 44 clones for PL1 and PL2. The frequency of scFv insert in-frame was 61% for PL1 and 84.1% for PL2. These percentages were in good correlation with the frequencies observed in the synthetic fragments. Three in 44 clones (6%) of PL1 and one in 44 (2.3%) of PL2 were chimeric clones with two plasmids detected by sequencing.

EXAMPLE 5: BINDING OF PRIMARY LIBRARIES TO PROTEIN-A

[142] PL1 and PL2 binding to Protein A was assessed by a direct ELISA (**Figure 7**). Both PLs bound Protein A, with PL2 showing a higher EC_{50} than PL1. A higher EC_{50} in PL2 can be explained by a higher quality of the synthetic fragments since 4-01/3-23 synthetic fragments exhibited 60% of in-frame sequences, whereas, the 3-20/3-23 fragments had only 50%. Comparison with the control scaffolds having the CDR-H3 of CNTO 888 and displayed as sFv fusion to pIII showed similar signal in intensity at saturation level and inflexion at dilutions 20 to 6 times lower for PL1 and PL2, respectively. This

observation was consistent with the previous observation that only a fraction of the virions displays fusions on their surface.

EXAMPLE 6: PROTEIN-A FILTERING AFTER HEAT SHOCK

[143] To set up the optimal conditions for selection of developable antibody fragments, two sets of experiments were performed: (1) Parent scaffolds unfolding at diverse temperatures followed by capture with Protein A and (2) unfolding of diversified scaffold variants followed by capture with Protein A. In the first set of experiments the unfolding process of the parent scaffolds was studied. In the second set of experiments it was assessed how diversification of the scaffolds affected their folding and stability.

EXAMPLE 6.1: SCAFFOLDS UNFOLDING

[144] The 3-20/3-23 and 4-01/3-23 control scFvs with the CDR-H3 of CNTO 888 displayed as fusion proteins on the phage surface were incubated for 10 min in a range of temperatures, starting at 40°C and increasing the temperature up to 80°C in steps of 5°C (**Figure 8**). The unfolding process was monitored with a direct Protein A ELISA as described in Example 5. The 3-20/3-23 control scFv started to unfold at 65°C. The 4-01/3-23 scFv started the unfolding process at 55°C. The Tm (temperature at which 50% of the scFvs are unfolded) of 3-20/3-23 scFv had a value of 75°C. The Tm of 4-01/3-23 scFv was 65°C.

[145] The unfolding kinetic at 60°C and 72°C of the 3-20/3-23 control scFv and 4-01/3-23 control scFv (**Figure 9**) demonstrated that the 3-20/3-23 control scFv remained folded at 60°C for up to one hour. At 72°C it unfolded slowly with approximately a 20% drop in the ELISA signal during the first 30 min. The unfolding kinetic of 4-01/3-23 scFv was significantly different. It unfolded slowly at 60 °C. At 72°C the unfolding process accelerated, with a drop in the ELISA signal of approximately 30% in the first 10 min.

[146] Therefore, this example demonstrates that: (i) the 3-20/3-23 and 4-01/3-23 control scFvs are stable at 60°C for at least one hour; and (ii) 3-20/3-23 and 4-01/3-23 scFvs can be incubated at 72°C for 10 min without significant unfolding and aggregation.

EXAMPLE 6.2: FOLDING & STABILITY OF DIVERSIFIED SCAFFOLDS

[147] The folding and stability of the diversified scaffolds were assessed by analyzing both Protein A binding and unfolding at high temperature on single clones taken randomly from the primary libraries. A

subset of 44 clones from PL1 and 44 from PL2 were studied. Protein A was restricted to in-frame clones in both libraries (**Figure 10** and **Figure 11**). Only 4 chimeric clones, seen by the superimposition of two sequencing trace and most probably deriving from a double transformation were found, 3 in PL1 and 1 in PL2. These clones were part of the experimental process but were not included in the statistics.

[148] Because the Protein A binding signal in the ELISA exhibited an obvious saturation effect, most likely due to saturation of the Protein A binding sites at the bottom of the plate, binding were expressed as percentage of the respective control scaffold signal, ranging from no binding (e.g. clone A6 in PL1 or D1 in PL2) to 100% (**Figure 10** to **Figure 13**). Protein A binding of single clones was defined by at least 10% of the signal of the control scaffold. The characteristics of the selected clones in each library are summarized below (Table 3).

TABLE 3

Library[1]	**PL1**		**PL2**	
	Percentage	**Clones**	**Percentage**	**Clones**
Single Clones	93.2%	41/44	97.7%	43/44
Double Transformants	6.8%	3/44	2.1%	1/44
Vector Background	0.0%	0/44	0.0%	0/44
In-Frame Clones	61.0%	25/41	83.7%	36/43
Protein A Binding[2]	58.5%	24/41	69.8%	30/43
Heat Shock Survival[3]	48.8%	20/41	55.8%	24/43

[149] Protein A binding of single phage results from multiple factors such as: (1) as proper folding of the Fv fragment, (2) concentration of displayed scFv and (3) scFv valency, up to 5 per M13 phage. It should be noted that the last two properties are linked to better expression levels, e.g. the higher the expression, the higher the valency. Therefore, Protein A binding is linked to more developable antibodies (better folded or better expressed or both). On average 24 out of 41 single clones in PL1 bound Protein A (58.5%) and 30 out of 43 single clones in PL2 (69.8%) (Table 3). Therefore, the vast majority of the assayed clones had a signal between 20% and 90% with respect to the control scaffolds, showing a good dynamic range to assess changes in Protein A binding after heat treatment.

[150] The same clones were treated for 10 min at 70°C and analyzed the residual Protein A binding signal (**Figure 12** and **Figure 13**). As shown in Example 6.1, the control scaffold scaffolds started to unfold significantly in this temperature range. A preliminary analysis on a few clones at different temperatures showed a good compromise between clones that lost most of their binding and clones resistant to the heat shock. To prevent overestimation of thermal resistance, it was considered only clones

with at least 25% binding to the control scaffold at 37°C. Again, survival was defined as having a residual binding to Protein A of at least 10% of the control scaffold. On average 20 out of 41 single clones in PL1 survived the heat treatment (48.8%) and 24 out of 43 single clones survived in PL2 (55.8%) (Table 3). This observation indicated that around ~50% of the PLs were not functional after incubation at 70°C for 10 min.

EXAMPLE 7: PROTEIN A FILTRATION OF THE PRIMARY LIBRARIES

[151] Twenty PCR tubes containing each one 100 µl of virions from PL1 at a concentration of 1.3 x 10^{13} virions/ml (1.3 x 10^{12} virions per tube), were incubated at said temperature of 70°C for a 10 min period in a PCR machine and cooled down for 30 min on ice. In the meantime, 2 ml of magnetic beads (BioMag® Streptavidin #84660-5 from Polysciences, Inc., Warrington, PA) were washed two times with TBS with Tween 20 0.1% (TBST) and incubated with 500 µg biotinylated Protein A (Life Technologies, Cat #29989) for 30 min with agitation. After 3 washes with TBST, the beads were aliquoted in 10 microfuge tubes, resuspended in 400 µl TBST with non-fat dry milk 5% w/v and blocked for 30 min at room temperature. One-hundred µl of heated library were added to each tube and incubated for 2 h at room temperature on a rocker.

[152] After 5 washes with TBST and 5 washes with TBS, the bound phage were successively eluted with Trypsin-EDTA (Life Technologies, Cat# 25200056, 500 µl per tube) and glycine 0.1 mM, pH 2.7 containing BSA 1 mg/ml (500 µl per tube) for 10 min at room temperature. After neutralization of the acid eluate with Tris 1 M, pH 8.0, both eluates were combined and used to transfect XL10 Gold cells (Agilent Technologies, San Diego, CA). Bacteria were grown overnight at 37°C in 2xYT medium supplemented with ampicillin and in the presence of glucose 1% w/v. The day after, plasmid DNA were prepared using a DNA MIDI kit preparation (Macherey Nagel, Germany).

[153] For PL2, only 10 tubes were processed in a similar way for a total of half the number of initial virions. Bacterial transfectants incubated no later than 55 min post transfection were counted by dilutions on agar plates supplemented with ampicillin and glucose. The filtrated library FL1 had a total of 7.2 x 10^9 primary transfectants and FL2 had 5.1 x 10^9 primary transfectants.

[154] To assess the quality of Protein A filtering, a sample of 44 ampicillin-resistant colonies was selected at random in each library. Analysis by colony PCR for the presence of scFv showed that all 88 clones had an insert at the expected size. Further analysis by Sanger sequencing found seven double transformants in each library with the superimposition of two scFv sequence traces and only one clone out of frame in each library, giving an observed frequency of in-frame clones of 94.7% and 97.3% for FL1 and FL2, respectively, after excluding the double transformants.

[155] Phage displaying scFv for the single clones were prepared as described in Example 6.2 and binding to Protein A was demonstrated by ELISA. COSTAR plates 3369 (Corning) were coated with Protein A (Sigma Aldrich, cat# P6031) at 4 μg/ml in TBS overnight at 4°C. After blocking with TBST and 5% w/v nonfat dry milk for one hour, virions (~2.6 x 10^{12} virions/ml) in TBST with 5% w/v nonfat dry milk were added to the wells and incubated for 2 h at 37°C. As a reference, virions derived from the parent scaffolds with the CDR-H3 of CNTO 888 cloned in the same vector were added and similarly incubated on the plate. Bound phage was detected with A4G1.6 monoclonal antibody (Antibody Design Labs, San Diego, CA) conjugated to HRP. Binding of the secondary antibody, a murine IgG1, to Protein A was blocked by polyclonal human IgG at 100 μg/ml added into the incubation buffer as described in **Figure 7**. The characteristic of the selected clones are summarized below (Table 4).

TABLE 4

Library	FL1		FL2	
	Percentage	Clones	Percentage	Clones
Single Clones	84.1%	37/44	84.1%	37/44
Double Transformants	15.9%	6/44	15.9%	7/44
Vector Background	0.0%	0/44	4.3%	0/44
In-Frame Clones	94.7%	36/38	97.3%	36/37
Frequency of Protein A Binding	89.5%	34/38	89.2%	33/37

[156] The percentage of clones binding Protein A with ≥ 10% Protein A binding with respect to the control scaffold was of 91.9% for FL1 and 89.2% for FL2 (**Figure 14** and **Figure 15**, and Table 4). The very high frequency of insert clones in both libraries together with a frequency of Protein A binders around 90%, almost 50% more than the frequencies observed in PLs (58.5% and 65.2% for PL1 and PL2, respectively), is a clear signature of the effectiveness of the Protein A filtration. Based on the infectivity of the phage displaying scFv, around 5% in the pADL-23/scFv/M13KO7 system, it was found retrospectively that the MOI was above 1 during the infection of the XL10 Gold cells, explaining *a posteriori* the high frequency of double transformants in the FLs. Since virions were not prepared from these libraries but just the plasmid DNA, this was not consequential on the quality of the SL libraries.

EXAMPLE 8: SECONDARY REPERTOIRES

[157] In this example, we will create secondary libraries by replacing the H3J fragments of the filtrated libraries exemplified in Example 7 by a pool of natural H3JK fragment derived from a large pool of human donors.

EXAMPLE 8.1: NATURAL H3J FRAGMENTS FROM HUMAN REPERTOIRES

[158] The natural H3J fragments were obtained from the PBMCs of 200 healthy donors, 100 females and 100 males under the age of 40 years. Each donor provided 5 x 10^6 cells and thus potentially yielded 1 x 10^6 unique H3J sequences. Therefore, the pool of 200 donors contained 2 x 10^8 potentially unique H3J sequences. Starting from the PBMCs, total RNA (tRNA) was individually isolated using Trizol (Invitrogen; Cat# 15596026 and 15596018). Pools of tRNAs from 10 donors were generated after determining the concentration by UV spectrophotometry and mixing the donor tRNAs in equal amounts to generate 20 tRNA pools.

[159] Each of the 20 pools were processed to isolate Messenger RNA (mRNA) using polyA Spin™ mRNA Isolation Kit (NEB, Cat #: S1560S) following the manufacturer instructions. The mRNA was used as template to generate cDNA by reverse transcription using OneTaq® RT-PCR Kit (NEB, Cat #: E5310S) and a poly-T oligonucleotide.

[160] To amplify the natural CDR-H all human VH genes in the FR3 region were aligned immediately before the CDR-H3. This region of the heavy chain exhibits highly conserved amino acids with two tyrosines at position 90 and 91 and a cysteine at position 92 (Chothia and Lesk, 1987). This conservation is reflected as well at the nucleotide level. Second, the alignment was filtered by considering only VH genes commonly found in circulating PBMCs, more precisely VH genes that had been found in at least 0.6% of the human circulating antibody repertoire (Glanville J et al., 2009). This particular set of antibodies has a highly conserved nucleotide sequence GACACGGCYGTGTATTACTGTGC (SEQ ID NO: 94) located at the FR-3 – CDR-H3 junction. There is polymorphism with a C or a T for the third nucleotide of the alanine codon at position 88, hence the Y (UIPAC code for C or T) symbol in this sequence. This motif is present in more than 95% of the VH sequences found in circulating PBMCs. This human VH conserved motif (CM) was used to amplify the H3J fragments from the cDNA of the donor pool.

[161] Double-stranded DNA containing the repertoire of natural H3J fragments was obtained by PCR using a universal forward primer annealing to the CM and three reverse primers designed to amplify 95% of the human CDR-H3 fragments in circulating PBMCs (SEQ ID Nos: 95-97).

[162] The quality of H3J fragments was assessed by cloning an aliquot of final pool into a TOPO vector (Life Technologies). Sanger sequencing of 30 clones indicated that all H3J fragments were different, with length variation resembling the human CDR-H3 repertoire. The region introduced by the

amplification primers for assembling the full scFvs and cloning into the vector matched 100% the expected nucleotide sequence.

EXAMPLE 8.2: CLONING OF A SECONDARY REPERTOIRE BY SEAMLESS ASSEMBLY

[163] The strategy for seamless assembly of the SLs is highlighted on **Figure 16**. The region corresponding to V_L, GS19 linker peptide and V_H just before the CM motif (**Figure 16**) was amplified by PCR from the DNA purified from the filtered primary libraries with the sfiFOR primer taken on the pelB leader-encoding sequence (SEQ ID NO: 98) and the ALT_hu3_23FR3_r primer annealing just before the CM motif (SEQ ID NO: 99). ALT_hu3_23FR3_r is extended on its 5'end by the sequence 5'-CACAGGTCTCG. This sequence contains a Bsa-I site, which, after digestion, creates a 4-base pair overhangs TGTC complementary to the first four nucleotides GACA of the CM motif.

[164] The second step of the scFv assembly involved a simultaneous digestion by BsaI and ligation by T4 DNA ligase at 37°C which joined the primary filtered fragments and the natural H3J fragments. The resulting full length scFv DNA was further amplified by PCR before cloning (**Figure 16**).

[165] Amplification of the primary filtered library DNA from library FL1 and FL2 with 17 cycles and an annealing temperature of 60°C using Phusion (New England Biolabs, MA) yielded single bands at the expected size (**Figure 17, Panel A**). For each library, natural H3J fragments (150 ng) were assembled with the primary filtered DNA fragments (600 ng) by simultaneous digestion with BsaI and ligation with T4 ligase for 4 h at 37°C. The natural H3J were in a slight excess and conversion of the filtered fragments was near 100% (**Figure 17, Panel B**). 60 ng of the joined products were further amplified by the nested primers padllib_s (SEQ ID NO: 101) and ALT_huH3J_r (SEQ ID NO: 102) in a final volume of 300 µl using Phusion, 20 cycles and an annealing temperature of 60°C (**Figure 17, Panel C**).

[166] The two semisynthetic fragments so generated were ligated into the pADL-23c phagemid vector. A total of 5 µg of ligated products was electroporated into electro-competent TG1 cells (800 ng/50 µl cells) as described in Example 1 for each secondary library. The number of primary transformants for the secondary libraries was 1.4×10^{10} cfu for the SL1 library and 1.1×10^{10} cfu for the SL2 library.

[167] The quality of each secondary library was assessed by Sanger sequencing of 44 clones picked randomly from a separate electroporation made with 10 times less DNA (80 ng/50 µl cells). The results are summarized in Table 5 below. As in Table 3, binding to Protein A is again defined as more than 10% binding relative to the respective control scaffold. We found 3 double transformants in each SL1 and SL2 libraries. These clones were excluded from the calculations. Finally, it was found that 35/41 (85.4%)

were in-frame for the SL1 secondary library and 35/41 (85.4%) for the SL2 secondary library. These values were ~10% lower than those of the filtered libraries indicating that the method of secondary library construction was able to retain most of the in-frame character of the filtrated libraries while adding natural H3J diversity in the heavy chain, thus compensating the loss of diversity during the step of filtration.

TABLE 5

Library	SL1		SL2	
	Percentage	Clones	Percentage	Clones
Single Clones	95.5%	42/44	95.5%	40/44
Double Transformants	6.8%	3/44	6.8%	3/44
Vector Background	0.0%	0/41	0%	0/40
In-Frame Clones	85.4%	35/41	85.4%	35/41
Frequency of Protein A Binding	82.9%	34/41	85.4%	35/41
Heat Shock Survival	68.3%	28/41	70.7%	29/41

EXAMPLES 8.3: FOLDING & STABILITY OF THE SECONDARY LIBRARIES

[168] The 44 clones of each secondary library were tested for Protein A binding and survival after a heat shock at 70°C for 10 min (**Figure 19** and **Figure 20**). All in-frame clones but one in the SL1 library gave strong binding to Protein A, giving a frequency of Protein A binding of 85.4% in that library. All in-frame clones in SL2 library were binding Protein A, giving a frequency of Protein A binding of 85.4% in that library. Therefore, not only in-frame clones that were selected by Protein A during the filtration remained in-frame but were for most of them were still binding Protein A after replacing the H3J sequences. Analysis of survival to the heat shock of the clones for clones having 25% or more binding to Protein A relative to the respective control scaffold prior to the heat shock (**Figure 20** and **Figure 21**) found 28/41 (68.3%) for SL1 library, 19.5% better than PL1, and 29/41 (70.7%) for SL2, 14.9% better than the corresponding PL1. The same analysis for all clones found that 30/41 or 73.1% of the SL1 and 27/40 or 67.5% of the SL2 were surviving the heat shock. This was a significant increase comparatively to the PLs with 39.9% for PL1 and 33.3% for PL2 (**Figure 22**).

EXAMPLE 9: ASSESSING DIVERSITY OF THE LIBRARIES BY NEXT GENERATION SEQUENCING

[169] To study the diversity along the filtration process and the construction of the secondary libraries, two amplicons were prepared from the phagemid DNA. Plasmid DNA were isolated using QIAGEN Plasmid Midi Kit (Cat No.: 12143) from the PL and SL bacterial cultures prior to helper phage superinfection and induction, and after overnight bacterial culture for the FL libraries. The DNA was used as a template to generate amplicons of approximately 300 bps. Amplicon 1 covered the V_L scaffolds (3-20 and 4-01) and was amplified with one forward primer for PL1 (SEQ ID 126) and PL2 (SEQ ID 127) and a reverse primer located in the J region (SEQ ID 128). This amplicon included the three V_L CDRs. Amplicon 2 covered the CDR-H2 and the H3J fragments and was amplified with one forward primer in FR1 region (SEQ ID 129) and one reverse primer located in the vector sequence immediately after the SfiI site (SEQ ID 130). The PCR reactions were performed as follows: 5 min start at 95°C followed by 10 cycles of 1 min at 95°C, 1 min at 67°C, 1 min at 72°C and terminated by a 10-min extension at 72°C. The PCR fragments were gel-purified using QIAquick PCR Purification Kit (Cat No.: 28104) and used as template to prepare the samples for NGS following the manufacturer instructions. The sequencing was performed on a Miseq platform from Illumina. FASTQ files were processed with the software AptaAnalyzer™ (AptaIT; Germany) using the BCR (B-cell receptors) functionality.

[170] The accepted output sequences were further curated with in-house Java scripts. In brief, for amplicon 1, sequences having insertions or deletions of two amino acids or more were discarded. For amplicon 2 sequences before FR-3 region having insertions or deletions of more than two amino acids were discarded and for CDR-H3, sequences not having the conserved cysteine H88 and tryptophan H102 were removed from the analysis as well.

EXAMPLE 9.1: NGS ANALYSIS OF THE PRIMARY LIBRARIES AND THE FILTRATION PROCESS

[171] Over half million sequences were obtained from PL1 amplicon 1 and close to a million and a half for FL1 amplicon 1 (Table 6). A clear difference toward curated sequences from 76% in PL1 to up 95% in FL1 was observed, in agreement with an increase of productive sequences as a result of the filtration process. For PL2 and FL2, the total number of sequences was higher than for PL1, with over a million and a half for PL2 and over two million sequences for FL2 amplicon 1 (Table 7). The number of curated sequences in both PL2 and FL2 was higher than PL1 (88%) and similar for FL1 (90%). This observation agreed with the observation that the quality of synthetic PL2 fragments was superior to that of PL1.

TABLE 6

Amplicon 1	PL1		FL1	
	Count	%	Count	%
Total sequences	533,828		1,447,032	
Accepted sequences	497,359	100.00	1,349,887	100.00
Curated Sequences	379,834	76.37	1,284,091	95.13
Unique sequences	222,465	44.73	483,262	35.80

TABLE 7

Amplicon 1	PL2		FL2	
	Count	%	Count	%
Total sequences	1,466,411		2,249,262	
Accepted sequences	1,341,533	100.00	2,172,872	100.00
Curated Sequences	1,285,500	95.82	1,984,567	91.33
Unique sequences	766,255	57.12	1,014,666	46.70

[172] The number of total sequences in amplicon 2 for PLs and FLs was similar with close to one million and a half sequences for each library (Table 8 and 9). The curated sequences were also similar, close to ~ 90%.

TABLE 8

Amplicon 2	PL1		FL1	
	Count	%	Count	%
Total sequences	1,757,878		1,509,376	
Accepted sequences	1,490,192	100.00	1,308,501	100.00
Curated unique sequences	695,605	46.68	584,346	44.66

TABLE 9

Amplicon 2	PL2		FL2	
	Count	%	Count	%

Total sequences	1,637,123		1,389,471	
Accepted sequences	1,330,040	100.00	1,242,319	100.00
Curated unique sequences	656,092	49.33	586,474	47.21

[173] In both amplicons, the percentage of unique sequences was on average 50% in the PLs, providing a large coverage of the primary diversity. This percentage decreased on average of 5.8% during the filtration process, pointing to a net loss of diversity after treatment at 70°C for 10 min and re-amplification of the libraries.

[174] Analysis of the frequency of the neutral IGDH fragments contained in amplicon 2 before and after the filtration process (Table 10) indicated that all the designed fragments were represented in the PLs. The frequency was within 1.3% of expected value of 6% of an even distribution (18 neutral fragments) and was similar in both PLs, suggesting no major bias due to the PLs preparation.

[175] After the filtration process the difference in frequency between PLs and FLs of some neutral IGDH fragments increased by more than 25% in both FLs with respect to the PLs, whereas others decreased by a similar frequency or more (Table 11). Both FL1 and FL2 showed a similar trend, although FL1 showed overall variations. Of particular interest was the IGDH fragment "VDIVATI", which decreased by close to 75% in PL1 and slightly over 50% in PL2. This fragment contains five hydrophobic amino acids out of 7, suggesting an unfavorable selection for hydrophobic residues during the filtration process. Of the same trend was the consistent decrease after the filtration process of the longest IGDH fragments, all rich in tyrosine residues, and increase of the most hydrophilic IGDH "GITGT" and "GTTGT"

TABLE 10

Neutral D sequence	**PL1 %**	**PL2 %**	**FL1 %**	**FL2 %**
GYSGYDY	6.84%	6.83%	8.29%	7.90%
GYSYGY	6.88%	6.86%	7.12%	6.66%
TTVT	5.45%	5.53%	4.87%	5.00%
YSGSYY	4.63%	5.28%	4.49%	5.08%
DYGDY	5.51%	5.59%	6.07%	6.11%
DYSNY	5.53%	5.62%	5.76%	5.70%
SIAAR	5.40%	5.32%	6.08%	6.17%
VQLER	4.38%	4.41%	3.22%	4.03%
YYDILTGYYN	4.73%	4.43%	2.97%	3.28%
YYYDSSGYYY	4.53%	4.20%	1.94%	2.05%

YYYGSGSYYN	4.82%	4.46%	3.66%	3.48%
EYSSSS	6.13%	6.05%	6.66%	6.48%
GITGT	5.83%	5.82%	7.99%	7.33%
GIVGAT	5.99%	6.07%	6.99%	6.87%
GTTGT	5.28%	5.35%	7.89%	6.80%
GYSSGY	6.09%	6.15%	7.78%	7.11%
LTG	5.40%	5.59%	6.29%	6.80%
VDIVATI	6.58%	6.43%	1.94%	3.14%

[176] The five IGHJ germline genes were also within 4% of their even distribution (20%) with a slight 5% increase of IGHJ3 and IGHJ6 at the expense of IGHJ1 and IGHJ4 after filtration (**Table 11**). Again, the IGHJ frequencies were similar in both libraries, consistent with the suggestion that no major bias resulted from the PLs preparation.

TABLE 11

JH	**PL1 %**	**PL2 %**	**FL1 %**	**FL2 %**
JH1	20.51%	19.90%	17.10%	17.40%
JH3	21.02%	22.03%	27.02%	25.96%
JH4	20.51%	19.87%	14.84%	15.16%
JH5	21.63%	21.31%	22.00%	21.93%
JH6	16.33%	16.90%	19.04%	19.55%
JH1	20.51%	19.90%	17.10%	17.40%

EXAMPLE 9.2: NGS ANALYSIS OF THE SECONDARY LIBRARIES

[177] Over one million primary sequences were obtained for both SL1 and SL2. The number of accepted sequences was around 80% in both SLs, thus providing a large sampling of the secondary libraries at ~850,000 and ~990,000 sequence depth for SL1 and SL2, respectively (**Table 12**). Around 60% of these sequences had unique natural H3J fragments in each library. The top 30 most frequent sequences in each library barely added up to 1% of the total accepted sequences, indicating that the

remaining 99% of all CDR-H3 sequences had only a few copies. Only half of the most prevalent 30 sequences were shared by the two libraries, pointing to the large diversity of CDR-H3 in the SLs.

TABLE 12

Amplicon 2	SL1		SL2	
	Count	%	Count	%
Total sequences	1,077,746	100.00	1,227,176	100.00
Accepted sequences	853,327	79.18	990,864	80.74
Unique sequences (Synthetic diversity and natural H3J fragments)	851,681	99.81	989,005	99.81
Unique CDR-H3 sequences	522,939	61.28	592,317	59.78

[178] The natural CDR-H3 length distribution followed a Gaussian curve typical of the human CDR-H3 sequences with an expected peak at 12 amino acids in length (Kabat's definition, Table 13). No difference was observed between SL1 and SL2, indicating that the pool of natural H3J fragments from the 200 donors was not biased towards a particular CDR-H3 length during the SL preparation.

TABLE 13

CDR-H3 length	SL1		SL2	
	Count	%	Count	%
<4	3660	0.4	4588	0.5
4	3260	0.4	4179	0.4
5	4193	0.5	5333	0.5
6	9366	1.1	11640	1.2
7	19401	2.3	23589	2.4
8	32294	3.8	39128	3.9
9	52001	6.1	61615	6.2
10	69709	8.2	82621	8.3
11	84753	9.9	99212	10.0
12	99841	11.7	116562	11.8
13	95688	11.2	109993	11.1
14	85577	10.0	99002	10.0
15	76779	9.0	87474	8.8
16	66070	7.7	75163	7.6
17	54963	6.4	62589	6.3
18	43447	5.1	49095	5.0
19	31482	3.7	35744	3.6

20	19775	2.3	21975	2.2
>20	1093	0.1	1322	0.1

[179] The distribution of the IGHJ1-5 segments (Table 14) followed the expected usage seen in human antibody sequences except for IGHJ6. The reported IGJH6 frequency is ~ 40%, followed by IGHJ4 with ~ 30% (Arnaout et al., 2011). In the SLs, IGHJ4 was the most prevalent joining region with ~40%, followed by the IGHJ6 with 20%. However, it was so designed during the RT-PCR amplification of the natural H3J fragments since the IGHJ6 encodes a stretch of five Y residues in the N-terminal region, which could lead to a potential destabilizing effect. Thus, by lowering the proportion of natural H3J fragments encoded by IGHJ6 it was expected to select for a higher number of developable antibodies.

TABLE 14

JH	SL1		SL2	
	Count	%	Count	%
IGHJ4*01	337668	40	392998	40
IGHJ6*01	174385	20	201522	20
IGHJ3*02	94067	11	108560	11
IGHJ5*02	78043	9	89640	9
IGHJ1*01	55861	7	65708	7
IGHJ2*01	42988	5	49448	5
IGHJ3*01	24237	3	28729	3
IGHJ6*03	23977	3	27576	3
IGHJ5*01	20613	2	24887	3
IGHJ6*04	1488	0	1796	0

[180] When the CDR-H2 diversity was analyzed together with the CDR-H3 diversity, more than 99% of all accepted sequences were unique (Table 10), in clear increase from the filtrated libraries and pointing to the exquisite diversity of the SLs, thus confirming at the NGS level the successful increase in sequence diversity achieved during the construction of the secondary libraries.

EXAMPLE 10: ASSESSING FUNCTIONALITY OF THE SECONDARY LIBRARIES USING TWO TARGET MODELS

[181] To assess the potential of the secondary libraries to produce specific and developable antibodies, selections were performed with two non-related target models: TNF and HSA.

EXAMPLE 10.1: SELECTION WITH TNF

[182] The selections with TNF were performed in solid phase following protocols well known to those skilled in art. In brief, one Immunotube (Thermofisher cat #: 341866) was coated with 4 ml of TNF at a concentration of 50 µg/ml in Carbonate Buffer ($NaHCO_3$ 50 mM, pH 9.6) for one hour at room temperature (RT). To avoid non-specific interactions, the Immunotube was blocked with MPBS (PBS + 3% w/v skim powder milk) for one additional hour at RT and washed three times with PBS. For the first round of selection, the SL1 and SL2 libraries (3 x 10^{13} phage) were mixed and diluted to 1 x 10^{13} virion/ml in MPBS and 4 ml were added to the TNF-coated Immunotube. The Immunotube was incubated one hour at RT with slow shaking and one additional hour standing at RT. The unbound phages were washed away 10 times with TPBS (PBS + 0.1% Tween ® 20) washes and 10 washes with PBS. Bound phages were eluted with 0.2 M glycine/HCl, pH 2.2 and used to infect exponentially growing TG1 cells ($OD_{600\ nm}$ = 0.4). Infected cells were grown overnight at 37°C in 2xYT-agar plates containing 100 µg/ml carbenicillin and 1% w/v glucose. The cells were scrapped from the plate with 2xYT medium, expanded in 2xYT medium supplemented with carbenicillin and 1% w/v glucose, and superinfected with M13KO7 helper phage. After exchanging the medium with fresh 2xYT without glucose, cells were incubated overnight in the presence of carbenicillin 100 µg/ml and kanamycin 50 µg/ml at 30°C and 250 rpm. Virions were then purified by PEG-precipitation and used as input at a concentration of 1 x 10^{12} virion/ml for the following round of selection as described above.

[183] After the fourth round of panning, soluble scFvs of 45 clones chosen at random were assayed in ELISA for binding to TNF and BSA. As reporter reagent, Protein A/HRP was used. 12 positive clones for TNF but not for BSA were sent for Sanger sequencing to determine unique clones. Two unique clones were expressed as soluble scFvs and re-tested for binding to the target and BSA. The confirmed clones were expressed in 100-mL culture and purified using HisTrap™ (Sigma; GE Catalogue # GE17-5255-01*).* The purified scFvs exhibited specific binding to the target in a direct ELISA as shown on **Figure 23** or clone E12.

EXAMPLE 10.2: SELECTION WITH HSA

[184] The selections with HSA were conducted following a similar procedure as for TNF with some modifications. That is, only SL2 was used for the selections and a COSTAR plate 3369 (Corning) instead of Immunotubes was used as solid phase. Ten wells were coated with 100 µl per well of HSA 4 µg/ml in PBS overnight at 4°C. After washing 3 times with PBS and blocking with 200 µl TPBS with BSA 3% w/v for one hour at 37°C, 100 µl of SL2 phage, 6.1 x 10^{11} cfu/ml, per well in TPBS with BSA 3% w/v were incubated at 37°C for 2 h. After washing 5 times with TPBS and 5 times with TBS, bound phage were eluted by 100 µl glycine/HCl pH 2.2 BSA 1 mg/ml for 10 min at room temperature. Eluted fractions were collected and neutralized with Tris/HCl pH 8.0 1 M and used to infect exponentially growing TG1 cells as for the TNF screening.

[185] After 3 consecutive rounds of selection, 48 clones were analyzed by direct phage ELISA in 96-well format for binding to HSA with BSA as a negative control. 40 clones out of 48 showed a strong and specific binding to HSA (hit rate of 83%). Sanger sequencing of the 40 positive clones yielded 3 unique scFv sequences showing specific and intense binding to HSA in a phage ELISA and no binding to BSA as a negative control.

EXAMPLE 11: SEQUENCES

[186] As mentioned throughout the aforementioned examples, several sequences were used in the construction of libraries as discussed in this application. Exemplary sequences to be used in the foregoing examples include those shown below, in Table 15.

TABLE 15

SEQ ID NO: 1	GAAATTGTGTTGACGCAGTCTCCAGGCACCCTGTCTTTGTCTCCAGGGGA ACGTGCCACCCTCTCCTGCCGTGCCAGTCAGAGTGTTAGCAGCAGCTACT TAGCCTGGTATCAGCAGAAACCTGGCCAGGCTCCCCGACTCCTCATCTAT GGTGCATCTAGCCGTGCCACTGGTATCCCAGACCGTTTCAGTGGCAGTGG GTCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATT TTGCAGTGTATTACTGTCAGCAGTATGGTAGCTCACCTCTGACGTTCGGC CAAGGTACCAAGGTGGAAATCAAA
SEQ ID NO: 2	GACATCGTGATGACCCAGTCTCCAGACTCCCTGGCTGTGTCTCTGGGCGA GCGTGCCACCATCAACTGCAAGTCCAGCCAGAGTGTTTTATACAGCTCCA ACAATAAGAACTACTTAGCTTGGTATCAGCAGAAACCAGGACAGCCTCCT

	AAGCTGCTCATTTACTGGGCATCTACCCGGGAATCCGGGGTCCCTGACCG ATTCAGTGGCAGCGGGTCTGGGACAGATTTCACTCTCACCATCAGCAGCC TGCAGGCTGAAGATGTGGCAGTTTATTACTGTCAGCAATATTATAGTACT CCTCTGACGTTCGGCCAAGGTACCAAGGTGGAAATCAAA
SEQ ID NO: 3	GAGGTGCAGCTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGT CCCTGCGACTCTCCTGTGCAGCCTCTGGATTCACCTTTAGCAGCTATGCCA TGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTGTCAGC TATTAGTGGTAGTGGTGGTAGCACATACTACGCAGACTCCGTGAAGGGCC GGTTCACCATCTCCCGTGACAATTCCAAGAACACGCTGTATCTGCAAATG AACAGCCTGCGTGCCGAGGACACGGCCGTGTATTACTGTGCGAAATACG ACGGTATCTACGGTGAACTGGACTTCTGGGGCCAAGGAACCCTGGTCACC GTTTCCTCA
SEQ ID NO: 4	GYSGYDYAEYFQHWGQGTLVTVSS
SEQ ID NO: 5	GYSYGYAEYFQHWGQGTLVTVSS
SEQ ID NO: 6	TTVTAEYFQHWGQGTLVTVSS
SEQ ID NO: 7	YSGSYYAEYFQHWGQGTLVTVSS
SEQ ID NO: 8	DYGDYAEYFQHWGQGTLVTVSS
SEQ ID NO: 9	DYSNYAEYFQHWGQGTLVTVSS
SEQ ID NO: 10	SIAARAEYFQHWGQGTLVTVSS
SEQ ID NO: 11	VQLERAEYFQHWGQGTLVTVSS
SEQ ID NO: 12	YYDILTGYYNAEYFQHWGQGTLVTVSS
SEQ ID NO: 13	YYYDSSGYYYAEYFQHWGQGTLVTVSS
SEQ ID NO: 14	YYYGSGSYYNAEYFQHWGQGTLVTVSS
SEQ ID NO: 15	EYSSSSAEYFQHWGQGTLVTVSS
SEQ ID NO: 16	GITGTAEYFQHWGQGTLVTVSS
SEQ ID NO: 17	GIVGATAEYFQHWGQGTLVTVSS

SEQ ID NO: 18	GTTGTAEYFQHWGQGTLVTVSS
SEQ ID NO: 19	GYSSGYAEYFQHWGQGTLVTVSS
SEQ ID NO: 20	LTGAEYFQHWGQGTLVTVSS
SEQ ID NO: 21	VDIVATIAEYFQHWGQGTLVTVSS
SEQ ID NO: 22	GYSGYDYDAFDVWGQGTMVTVSS
SEQ ID NO: 23	GYSYGYDAFDVWGQGTMVTVSS
SEQ ID NO: 24	TTVTDAFDVWGQGTMVTVSS
SEQ ID NO: 25	YSGSYYDAFDVWGQGTMVTVSS
SEQ ID NO: 26	DYGDYDAFDVWGQGTMVTVSS
SEQ ID NO: 27	DYSNYDAFDVWGQGTMVTVSS
SEQ ID NO: 28	SIAARDAFDVWGQGTMVTVSS
SEQ ID NO: 29	VQLERDAFDVWGQGTMVTVSS
SEQ ID NO: 30	YYDILTGYYNDAFDVWGQGTMVTVSS
SEQ ID NO: 31	YYYDSSGYYYDAFDVWGQGTMVTVSS
SEQ ID NO: 32	YYYGSGSYYNDAFDVWGQGTMVTVSS
SEQ ID NO: 33	EYSSSSDAFDVWGQGTMVTVSS
SEQ ID NO: 34	GITGTDAFDVWGQGTMVTVSS
SEQ ID NO: 35	GIVGATDAFDVWGQGTMVTVSS
SEQ ID NO: 36	GTTGTDAFDVWGQGTMVTVSS
SEQ ID NO: 37	GYSSGYDAFDVWGQGTMVTVSS
SEQ ID NO: 38	LTGDAFDVWGQGTMVTVSS
SEQ ID NO: 39	VDIVATIDAFDVWGQGTMVTVSS
SEQ ID NO: 40	GYSGYDYYFDYWGQGTLVTVSS

SEQ ID NO: 41	GYSYGYYFDYWGQGTLVTVSS
SEQ ID NO: 42	TTVTYFDYWGQGTLVTVSS
SEQ ID NO: 43	YSGSYYYFDYWGQGTLVTVSS
SEQ ID NO: 44	DYGDYYFDYWGQGTLVTVSS
SEQ ID NO: 45	DYSNYYFDYWGQGTLVTVSS
SEQ ID NO: 46	SIAARYFDYWGQGTLVTVSS
SEQ ID NO: 47	VQLERYFDYWGQGTLVTVSS
SEQ ID NO: 48	YYDILTGYYNYFDYWGQGTLVTVSS
SEQ ID NO: 49	YYYDSSGYYYYFDYWGQGTLVTVSS
SEQ ID NO: 50	YYYGSGSYYNYFDYWGQGTLVTVSS
SEQ ID NO: 51	EYSSSSYFDYWGQGTLVTVSS
SEQ ID NO: 52	GITGTYFDYWGQGTLVTVSS
SEQ ID NO: 53	GIVGATYFDYWGQGTLVTVSS
SEQ ID NO: 54	GTTGTYFDYWGQGTLVTVSS
SEQ ID NO: 55	GYSSGYYFDYWGQGTLVTVSS
SEQ ID NO: 56	LTGYFDYWGQGTLVTVSS
SEQ ID NO: 57	VDIVATIYFDYWGQGTLVTVSS
SEQ ID NO: 58	GYSGYDYNWFDSWGQGTLVTVSS
SEQ ID NO: 59	GYSYGYNWFDSWGQGTLVTVSS
SEQ ID NO: 60	TTVTNWFDSWGQGTLVTVSS
SEQ ID NO: 61	YSGSYYNWFDSWGQGTLVTVSS
SEQ ID NO: 62	DYGDYNWFDSWGQGTLVTVSS
SEQ ID NO: 63	DYSNYNWFDSWGQGTLVTVSS

SEQ ID NO: 64	SIAARNWFDSWGQGTLVTVSS
SEQ ID NO: 65	VQLERNWFDSWGQGTLVTVSS
SEQ ID NO: 66	YYDILTGYYNNWFDSWGQGTLVTVSS
SEQ ID NO: 67	YYYDSSGYYYNWFDSWGQGTLVTVSS
SEQ ID NO: 68	YYYGSGSYYNNWFDSWGQGTLVTVSS
SEQ ID NO: 69	EYSSSSNWFDSWGQGTLVTVSS
SEQ ID NO: 70	GITGTNWFDSWGQGTLVTVSS
SEQ ID NO: 71	GIVGATNWFDSWGQGTLVTVSS
SEQ ID NO: 72	GTTGTNWFDSWGQGTLVTVSS
SEQ ID NO: 73	GYSSGYNWFDSWGQGTLVTVSS
SEQ ID NO: 74	LTGNWFDSWGQGTLVTVSS
SEQ ID NO: 75	VDIVATINWFDSWGQGTLVTVSS
SEQ ID NO: 76	GYSGYDYYGMDVWGQGTTVTVSS
SEQ ID NO: 77	GYSYGYYGMDVWGQGTTVTVSS
SEQ ID NO: 78	TTVTYGMDVWGQGTTVTVSS
SEQ ID NO: 79	YSGSYYYGMDVWGQGTTVTVSS
SEQ ID NO: 80	DYGDYYGMDVWGQGTTVTVSS
SEQ ID NO: 81	DYSNYYGMDVWGQGTTVTVSS
SEQ ID NO: 82	SIAARYGMDVWGQGTTVTVSS
SEQ ID NO: 83	VQLERYGMDVWGQGTTVTVSS
SEQ ID NO: 84	YYDILTGYYNYGMDVWGQGTTVTVSS
SEQ ID NO: 85	YYYDSSGYYYYGMDVWGQGTTVTVSS
SEQ ID NO: 86	YYYGSGSYYNYGMDVWGQGTTVTVSS

SEQ ID NO: 87	EYSSSSYGMDVWGQGTTVTVSS
SEQ ID NO: 88	GITGTYGMDVWGQGTTVTVSS
SEQ ID NO: 89	GIVGATYGMDVWGQGTTVTVSS
SEQ ID NO: 90	GTTGTYGMDVWGQGTTVTVSS
SEQ ID NO: 91	GYSSGYYGMDVWGQGTTVTVSS
SEQ ID NO: 92	LTGYGMDVWGQGTTVTVSS
SEQ ID NO: 93	VDIVATIYGMDVWGQGTTVTVSS
SEQ ID NO: 94	GACACGGCYGTGTATTACTGTGC
SEQ ID NO: 95	TGTTGGCCTCCCGGGCCTGAAGAGACGGTGACCATTGTCC
SEQ ID NO: 96	TGTTGGCCTCCCGGGCCTGAGGAGACGGTGACCGTGGTC
SEQ ID NO: 97	TGTTGGCCTCCCGGGCCTGAGGAGACRGTGACCAGGGT
SEQ ID NO: 98	CGCTGGATTGTTATTACTCGCG
SEQ ID NO: 99	CACAGGTCTCGTGTCCTCGGCACGCAGGCTGTTCATTTG
SEQ ID NO: 100	GTGTGGTCTCGGACACGGCYGTGTATTACTGTGC
SEQ ID NO: 101	TCGCGGCCCAGCCGGCCATG
SEQ ID NO: 102	GTGTTGGCCTCCCGGGCCTG
SEQ ID NO: 103	CAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCT CAGTGAAGGTCTCCTGCAAGGCTTCTGGATACACCTTCACCGGCTACTAT ATGCACTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGAC GGATCAACCCTAACAGTGGTGGCACAAACTATGCACAGAAGTTTCAGGG CAGGGTCACCAGTACCAGGGACACGTCCATCAGCACAGCCTACATGGAG CTGAGCAGGCTGAGATCTGACGACACGGTCGTGTATTACTGTGCGAGAG A
SEQ ID NO: 104	CAGGTTCAGCTGGTGCAGTCTGGAGCTGAGGTGAAGAAGCCTGGGGCCT CAGTGAAGGTCTCCTGCAAGGCTTCTGGTTACACCTTTACCAGCTATGGT ATCAGCTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGAT

	GGATCAGCGCTTACAATGGTAACACAAACTATGCACAGAAGCTCCAGGG CAGAGTCACCATGACCACAGACACATCCACGAGCACAGCCTACATGGAG CTGAGGAGCCTGAGATCTGACGACACGGCCGTGTATTACTGTGCGAGAG A
SEQ ID NO: 105	CAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCT CAGTGAAGGTTTCCTGCAAGGCATCTGGATACACCTTCACCAGCTACTAT ATGCACTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGAA TAATCAACCCTAGTGGTGGTAGCACAAGCTACGCACAGAAGTTCCAGGG CAGAGTCACCATGACCAGGGACACGTCCACGAGCACAGTCTACATGGAG CTGAGCAGCCTGAGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAG A
SEQ ID NO: 106	CAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGTCCT CGGTGAAGGTCTCCTGCAAGGCTTCTGGAGGCACCTTCAGCAGCTATGCT ATCAGCTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGAG GGATCATCCCTATCTTTGGTACAGCAAACTACGCACAGAAGTTCCAGGGC AGAGTCACGATTACCGCGGACGAATCCACGAGCACAGCCTACATGGAGC TGAGCAGCCTGAGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA
SEQ ID NO: 107	GAGGTGCAGCTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGT CCCTGAGACTCTCCTGTGCAGCCTCTGGATTCACCTTTAGCAGCTATGCCA TGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCAGC TATTAGTGGTAGTGGTGGTAGCACATACTACGCAGACTCCGTGAAGGGCC GGTTCACCATCTCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATG AACAGCCTGAGAGCCGAGGACACGGCCGTATATTACTGTGCGAAAGA
SEQ ID NO: 108	CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGT CCCTGAGACTCTCCTGTGCAGCCTCTGGATTCACCTTCAGTAGCTATGCTA TGCACTGGGTCCGCCAGGCTCCAGGCAAGGGGCTAGAGTGGGTGGCAGT TATATCATATGATGGAAGTAATAAATACTACGCAGACTCCGTGAAGGGCC GATTCACCATCTCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATG AACAGCCTGAGAGCTGAGGACACGGCTGTGTATTACTGTGCGAGAGA
SEQ ID NO: 109	GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGT CCCTGAGACTCTCCTGTGCAGCCTCTGGATTCACCTTCAGTAGCTATAGC

	ATGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTTCAT ACATTAGTAGTAGTAGTAGTACCATATACTACGCAGACTCTGTGAAGGGC CGATTCACCATCTCCAGAGACAATGCCAAGAACTCACTGTATCTGCAAAT GAACAGCCTGAGAGCCGAGGACACGGCTGTGTATTACTGTGCGAGAGA
SEQ ID NO: 110	CAGGTGCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGA CCCTGTCCCTCACCTGCGCTGTCTATGGTGGGTCCTTCAGTGGTTACTACT GGAGCTGGATCCGCCAGCCCCCAGGGAAGGGGCTGGAGTGGATTGGGGA AATCAATCATAGTGGAAGCACCAACTACAACCCGTCCCTCAAGAGTCGA GTCACCATATCAGTAGACACGTCCAAGAACCAGTTCTCCCTGAAGCTGAG CTCTGTGACCGCCGCGGACACGGCTGTGTATTACTGTGCGAGAGG
SEQ ID NO: 111	CAGGTGCAGCTGCAGGAGTCGGGCCCAGGACTGGTGAAGCCTTCGGAGA CCCTGTCCCTCACCTGCACTGTCTCTGGTGGCTCCATCAGTAGTTACTACT GGAGCTGGATCCGGCAGCCCCCAGGGAAGGGACTGGAGTGGATTGGGTA TATCTATTACAGTGGGAGCACCAACTACAACCCCTCCCTCAAGAGTCGAG TCACCATATCAGTAGACACGTCCAAGAACCAGTTCTCCCTGAAGCTGAGC TCTGTGACCGCTGCGGACACGGCCGTGTATTACTGTGCGAGAGA
SEQ ID NO: 112	GAGGTGCAGCTGGTGCAGTCTGGAGCAGAGGTGAAAAAGCCCGGGGAGT CTCTGAAGATCTCCTGTAAGGGTTCTGGATACAGCTTTACCAGCTACTGG ATCGGCTGGGTGCGCCAGATGCCCGGGAAAGGCCTGGAGTGGATGGGGA TCATCTATCCTGGTGACTCTGATACCAGATACAGCCCGTCCTTCCAAGGC CAGGTCACCATCTCAGCCGACAAGTCCATCAGCACCGCCTACCTGCAGTG GAGCAGCCTGAAGGCCTCGGACACCGCCATGTATTACTGTGCGAGACA
SEQ ID NO: 113	CAGGTACAGCTGCAGCAGTCAGGTCCAGGACTGGTGAAGCCCTCGCAGA CCCTCTCACTCACCTGTGCCATCTCCGGGGACAGTGTCTCTAGCAACAGT GCTGCTTGGAACTGGATCAGGCAGTCCCCATCGAGAGGCCTTGAGTGGCT GGGAAGGACATACTACAGGTCCAAGTGGTATAATGATTATGCAGTATCTG TGAAAAGTCGAATAACCATCAACCCAGACACATCCAAGAACCAGTTCTC CCTGCAGCTGAACTCTGTGACTCCCGAGGACACGGCTGTGTATTACTGTG CAAGAGA
SEQ ID NO: 114	GACATCCAGATGACCCAGTCTCCTTCCACCCTGTCTGCATCTGTAGGAGA CAGAGTCACCATCACTTGCCGGGCCAGTCAGAGTATTAGTAGCTGGTTGG

	CCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGAT GCCTCCAGTTTGGAAAGTGGGGTCCCATCAAGGTTCAGCGGCAGTGGATC TGGGACAGAATTCACTCTCACCATCAGCAGCCTGCAGCCTGATGATTTTG CAACTTATTACTGCCAACAGTATAATAGTTATTCTCC
SEQ ID NO: 115	GACATCCAGATGACCCAGTCTCCATCTTCCGTGTCTGCATCTGTAGGAGA CAGAGTCACCATCACTTGTCGGGCGAGTCAGGGTATTAGCAGCTGGTTAG CCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCT GCATCCAGTTTGCAAAGTGGGGTCCCATCAAGGTTCAGCGGCAGTGGATC TGGGACAGATTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAGATTTTG CAACTTACTATTGTCAACAGGCTAACAGTTTCCCTCC
SEQ ID NO: 116	GACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGA CAGAGTCACCATCACTTGCCAGGCGAGTCAGGACATTAGCAACTATTTAA ATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTACGAT GCATCCAATTTGGAAACAGGGGTCCCATCAAGGTTCAGTGGAAGTGGAT CTGGGACAGATTTTACTTTCACCATCAGCAGCCTGCAGCCTGAAGATATT GCAACATATTACTGTCAACAGTATGATAATCTCCCTCC
SEQ ID NO: 117	GATATTGTGATGACTCAGTCTCCACTCTCCCTGCCCGTCACCCCTGGAGA GCCGGCCTCCATCTCCTGCAGGTCTAGTCAGAGCCTCCTGCATAGTAATG GATACAACTATTTGGATTGGTACCTGCAGAAGCCAGGGCAGTCTCCACAG CTCCTGATCTATTTGGGTTCTAATCGGGCCTCCGGGGTCCCTGACAGGTTC AGTGGCAGTGGATCAGGCACAGATTTTACACTGAAAATCAGCAGAGTGG AGGCTGAGGATGTTGGGGTTTATTACTGCATGCAAGCTCTACAAACTCCT CC
SEQ ID NO: 118	GACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGA CAGAGTCACCATCACTTGCCGGGCAAGTCAGAGCATTAGCAGCTATTTAA ATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCT GCATCCAGTTTGCAAAGTGGGGTCCCATCAAGGTTCAGTGGCAGTGGATC TGGGACAGATTTCACTCTCACCATCAGCAGTCTGCAACCTGAAGATTTTG CAACTTACTACTGTCAACAGAGTTACAGTACCCCTCC
SEQ ID NO: 119	GAAATTGTGTTGACACAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGGA AAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAGCAGCTACTTAG

	CCTGGTACCAACAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGAT GCATCCAACAGGGCCACTGGCATCCCAGCCAGGTTCAGTGGCAGTGGGT CTGGGACAGACTTCACTCTCACCATCAGCAGCCTAGAGCCTGAAGATTTT GCAGTTTATTACTGTCAGCAGCGTAGCAACTGGCCTCC
SEQ ID NO: 120	GAAATAGTGATGACGCAGTCTCCAGCCACCCTGTCTGTGTCTCCAGGGGA AAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAGCAGCAACTTAG CCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGGT GCATCCACCAGGGCCACTGGTATCCCAGCCAGGTTCAGTGGCAGTGGGTC TGGGACAGAGTTCACTCTCACCATCAGCAGCCTGCAGTCTGAAGATTTTG CAGTTTATTACTGTCAGCAGTATAATAACTGGCCTCC
SEQ ID NO: 121	GAAATTGTGTTGACGCAGTCTCCAGGCACCCTGTCTTTGTCTCCAGGGGA AAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAGCAGCAGCTACT TAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTAT GGTGCATCCAGCAGGGCCACTGGCATCCCAGACAGGTTCAGTGGCAGTG GGTCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGAT TTTGCAGTGTATTACTGTCAGCAGTATGGTAGCTCACCTCC
SEQ ID NO: 122	GACATCGTGATGACCCAGTCTCCAGACTCCCTGGCTGTGTCTCTGGGCGA GAGGGCCACCATCAACTGCAAGTCCAGCCAGAGTGTTTTATACAGCTCCA ACAATAAGAACTACTTAGCTTGGTACCAGCAGAAACCAGGACAGCCTCC TAAGCTGCTCATTTACTGGGCATCTACCCGGGAATCCGGGGTCCCTGACC GATTCAGTGGCAGCGGGTCTGGGACAGATTTCACTCTCACCATCAGCAGC CTGCAGGCTGAAGATGTGGCAGTTTATTACTGTCAGCAATATTATAGTAC TCCTCC
SEQ ID NO: 123	TCTTCTGAGCTGACTCAGGACCCTGCTGTGTCTGTGGCCTTGGGACAGAC AGTCAGGATCACATGCCAAGGAGACAGCCTCAGAAGCTATTATGCAAGC TGGTACCAGCAGAAGCCAGGACAGGCCCCTGTACTTGTCATCTATGGTAA AAACAACCGGCCCTCAGGGATCCCAGACCGATTCTCTGGCTCCAGCTCAG GAAACACAGCTTCCTTGACCATCACTGGGGCTCAGGCGGAAGATGAGGC TGACTATTACTGTAACTCCCGGGACAGCAGTGGTAACCATCT
SEQ ID NO: 124	AATTTTATGCTGACTCAGCCCCACTCTGTGTCGGAGTCTCCGGGGAAGAC GGTAACCATCTCCTGCACCCGCAGCAGTGGCAGCATTGCCAGCAACTATG

	TGCAGTGGTACCAGCAGCGCCCGGGCAGTTCCCCCACCACTGTGATCTAT GAGGATAACCAAAGACCCTCTGGGGTCCCTGATCGGTTCTCTGGCTCCAT CGACAGCTCCTCCAACTCTGCCTCCCTCACCATCTCTGGACTGAAGACTG AGGACGAGGCTGACTACTACTGTCAGTCTTATGATAGCAGCAATCA
SEQ ID NO: 125	TCCTATGTGCTGACTCAGCCACCCTCAGTGTCAGTGGCCCCAGGAAAGAC GGCCAGGATTACCTGTGGGGGAAACAACATTGGAAGTAAAAGTGTGCAC TGGTACCAGCAGAAGCCAGGCCAGGCCCCTGTGCTGGTCATCTATTATGA TAGCGACCGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAACTCTG GGAACACGGCCACCCTGACCATCAGCAGGGTCGAAGCCGGGGATGAGGC CGACTATTACTGTCAGGTGTGGGACAGTAGTAGTGATCATCC
SEQ ID NO: 126	TCTTTCCCTACACGACGCTCTTCCGATCTTGGCAGAAATTGTGTTGACGCA G
SEQ ID NO: 127	TCTTTCCCTACACGACGCTCTTCCGATCTGACTCCCTGGCTGTGTCTCT
SEQ ID NO: 128	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCCTTGGTACCTTGGCCG AAC
SEQ ID NO: 129	TCTTTCCCTACACGACGCTCTTCCGATCTGTGCAGCCTCTGGATTCACC
SEQ ID NO: 130	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGTGGTGATGGTGTTGGC CTC

[187] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above can be used in various combinations.

REFERENCES

[188] Almagro, J.C., Teplyakov, A., Luo, J., Sweet, R.W., Kodangattil, S., Hernandez-Guzman, F., et al. (2014). Second antibody modeling assessment (AMA-II). *Proteins 82(*8**),** 1553-1562. doi: 10.1002/prot.24567.

[189] Arnaout R, Lee W, Cahill P, Honan T, Sparrow T, Weiand M, et al. High-resolution description of antibody heavy-chain repertoires in humans. PLoS One 2011; 6:e22365.

[190] Beerli, R.R., Bauer, M., Buser, R.B., Gwerder, M., Muntwiler, S., Maurer, P., et al. (2008). Isolation of human monoclonal antibodies by mammalian cell display. *Proc Natl Acad Sci U S A* 105(38)**,** 14336-14341. doi: 10.1073/pnas.0805942105.

[191] Bethea, D., Wu, S.J., Luo, J., Hyun, L., Lacy, E.R., Teplyakov, A., et al. (2012). Mechanisms of self-association of a human monoclonal antibody CNTO607. *Protein Eng Des Sel* 25(10), 531-537. doi: 10.1093/protein/gzs047.

[192] Bird, R.E., Hardman, K.D., Jacobson, J.W., Johnson, S., Kaufman, B.M., Lee, S.M., et al. (1988). Single-chain antigen-binding proteins. *Science* 242(4877), 423-426.

[193] Cherf, G.M., and Cochran, J.R. (2015). Applications of Yeast Surface Display for Protein Engineering. *Methods Mol Biol* 1319, 155-175. doi: 10.1007/978-1-4939-2748-7_8.

[194] Ching, K.H., Collarini, E.J., Abdiche, Y.N., Bedinger, D., Pedersen, D., Izquierdo, S., et al. (2017). Chickens with humanized immunoglobulin genes generate antibodies with high affinity and broad epitope coverage to conserved targets. *MAbs*, 1-10. doi: 10.1080/19420862.2017.1386825.

[195] Chothia, C., and Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 196(4), 901-917.

[196] de Haard, H.J., van Neer, N., Reurs, A., Hufton, S.E., Roovers, R.C., Henderikx, P., et al. (1999). A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *J Biol Chem* 274(26), 18218-18230.

[197] Emmons, C., and Hunsicker, L.G. (1987). Muromonab-CD3 (Orthoclone OKT3): the first monoclonal antibody approved for therapeutic use. *Iowa Med* 77(2), 78-82.

[198] Famm, K., Hansen, L., Christ, D., and Winter, G. (2008). Thermodynamically stable aggregation-resistant antibody domains through directed evolution. *J Mol Biol* 376(4), 926-931. doi: 10.1016/j.jmb.2007.10.075.

[199] Finlay, W.J., and Almagro, J.C. (2012). Natural and man-made V-gene repertoires for antibody discovery. *Front Immunol* 3, 342. doi: 10.3389/fimmu.2012.00342.

[200] Francisco, J.A., Campbell, R., Iverson, B.L., and Georgiou, G. (1993). Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. *Proc Natl Acad Sci U S A* 90(22), 10444-10448.

[201] Gilliland, G.L., Luo, J., Vafa, O., and Almagro, J.C. (2012). Leveraging SBDD in protein therapeutic development: antibody engineering. *Methods Mol Biol* 841, 321-349. doi: 10.1007/978-1-61779-520-6_14.

[202] Glanville J., Zhai W., Berka J., Telman D., Huerta G., Mehta G.R., Ni I., Mei L., Sundar P.D., Day G.M., Cox D., Rajpal A., and Pons J. (2009). Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci U S A*. 106(48):20216-2 .

[203] Green, L.L., Hardy, M.C., Maynard-Currie, C.E., Tsuda, H., Louie, D.M., Mendez, M.J., et al. (1994). Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nat Genet* 7(1), 13-21. doi: 10.1038/ng0594-13.

[204] Green, L.L., and Jakobovits, A. (1998). Regulation of B cell development by variable gene complexity in mice reconstituted with human immunoglobulin yeast artificial chromosomes. *J Exp Med* 188(3), 483-495.

[205] Griffiths, A.D., Williams, S.C., Hartley, O., Tomlinson, I.M., Waterhouse, P., Crosby, W.L., et al. (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J* 13(14), 3245-3260.

[206] Hanes, J., and Pluckthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A* 94(10), 4937-4942.

[207] Hoet, R.M., Cohen, E.H., Kent, R.B., Rookey, K., Schoonbroodt, S., Hogan, S., et al. (2005). Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. *Nat Biotechnol* 23(3), 344-348. doi: 10.1038/nbt1067.

[208] Hoogenboom, H.R. (2005). Selecting and screening recombinant antibody libraries. *Nat Biotechnol* 23(9), 1105-1116. doi: 10.1038/nbt1126.

[209] Hussack, G., Hirama, T., Ding, W., Mackenzie, R., and Tanha, J. (2011). Engineered single-domain antibodies with high protease resistance and thermal stability. *PLoS One* 6(11), e28218. doi: 10.1371/journal.pone.0028218.

[210] Jespers, L., Schon, O., Famm, K., and Winter, G. (2004). Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat Biotechnol* 22(9), 1161-1165. doi: 10.1038/nbt1000.

[211] Jones, P.T., Dear, P.H., Foote, J., Neuberger, M.S., and Winter, G. (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* 321(6069), 522-525.

[212] Kelly, R.L., Geoghegan, J.C., Feldman, J., Jain, T., Kauke, M., Le, D., et al. (2017). Chaperone proteins as single component reagents to assess antibody nonspecificity. *MAbs* 9(7), 1036-1040. doi: 10.1080/19420862.2017.1356529.

[213] Kimball, J.A., Norman, D.J., Shield, C.F., Schroeder, T.J., Lisi, P., Garovoy, M., et al. (1993). OKT3 antibody response study (OARS): a multicenter comparative study. *Transplant Proc* 25(1 Pt 1), 558-560.

[214] Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., et al. (2000). Fully

synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol* 296(1), 57-86. doi: 10.1006/jmbi.1999.3444.

[215] Köhler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256, 495-497.

[216] Langone, J.J. (1982). Protein A of Staphylococcus aureus and related immunoglobulin receptors produced by streptococci and pneumonococci. *Adv Immunol* 32, 157-252.

[217] Lonberg, N., Taylor, L.D., Harding, F.A., Trounstine, M., Higgins, K.M., Schramm, S.R., et al. (1994). Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* 368(6474), 856-859. doi: 10.1038/368856a0.

[218] Ma, B., Osborn, M.J., Avis, S., Ouisse, L.H., Menoret, S., Anegon, I., et al. (2013). Human antibody expression in transgenic rats: comparison of chimeric IgH loci with human VH, D and JH but bearing different rat C-gene regions. *J Immunol Methods* 400-401, 78-86. doi: 10.1016/j.jim.2013.10.007.

[219] Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D., and Winter, G. (1991). By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J Mol Biol* 222(3), 581-597.

[220] Matochko, W.L., Cory Li, S., Tang, S.K., and Derda, R. (2014). Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res* 42(3), 1784-1798. doi: 10.1093/nar/gkt1104.

[221] McCafferty, J., Griffiths, A.D., Winter, G., and Chiswell, D.J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348(6301), 552-554.

[222] Morrison, S.L., Johnson, M.J., Herzenberg, L.A., and Oi, V.T. (1984). Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc Natl Acad Sci U S A* 81(21), 6851-6855.

[223] Nechansky, A. (2010). HAHA--nothing to laugh about. Measuring the immunogenicity (human anti-human antibody response) induced by humanized monoclonal antibodies applying ELISA and SPR technology. *J Pharm Biomed Anal* 51(1), 252-254. doi: 10.1016/j.jpba.2009.07.013.

[224] Nilson, B.H., Solomon, A., Bjorck, L., and Akerstrom, B. (1992). Protein L from Peptostreptococcus magnus binds to the kappa light chain variable domain. *J Biol Chem* 267(4), 2234-2239.

[225] Obmolova, G., Teplyakov, A., Malia, T.J., Grygiel, T.L., Sweet, R., Snyder, L.A., et al. (2012). Structural basis for high selectivity of anti-CCL2 neutralizing antibody CNTO 888. *Mol Immunol*

51(2), 227-233. doi: 10.1016/j.molimm.2012.03.022.

[226] Perelson, A.S., and Oster, G.F. (1979). Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J Theor Biol* 81(4), 645-670.

[227] Pruzina, S., Williams, G.T., Kaneva, G., Davies, S.L., Martin-Lopez, A., Bruggemann, M., et al. (2011). Human monoclonal antibodies to HIV-1 gp140 from mice bearing YAC-based human immunoglobulin transloci. *Protein Eng Des Sel* 24(10), 791-799. doi: 10.1093/protein/gzr038.

[228] Raghunathan, G., Smart, J., Williams, J., and Almagro, J.C. (2012). Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens. *J Mol Recognit* 25(3), 103-113. doi: 10.1002/jmr.2158.

[229] Reichert, J.M. (2017). Antibodies to watch in 2017. *MAbs* 9(2), 167-181. doi: 10.1080/19420862.2016.1269580.

[230] Rouet, R., Lowe, D., and Christ, D. (2014). Stability engineering of the human antibody repertoire. *FEBS Lett* 588(2), 269-277. doi: 10.1016/j.febslet.2013.11.029.

[231] Shawler, D.L., Bartholomew, R.M., Smith, L.M., and Dillman, R.O. (1985). Human immune response to multiple injections of murine monoclonal IgG. *J Immunol* 135(2), 1530-1535.

[232] Shi, L., Wheeler, J.C., Sweet, R.W., Lu, J., Luo, J., Tornetta, M., et al. (2010). De novo selection of high-affinity antibodies from synthetic fab libraries displayed on phage as pIX fusion proteins. *J Mol Biol* 397(2), 385-396. doi: 10.1016/j.jmb.2010.01.034.

[233] Smith, S.L. (1996). Ten years of Orthoclone OKT3 (muromonab-CD3): a review. *J Transpl Coord* 6(3), 109-119; quiz 120-101.

[234] Sondek, J., and Shortle, D. (1992). A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proc Natl Acad Sci U S A* 89(8), 3581-3585.

[235] Strohl, W.R. (2017). Current progress in innovative engineered antibodies. *Protein Cell*. doi: 10.1007/s13238-017-0457-8.

[236] Teplyakov, A., Obmolova, G., Malia, T.J., Luo, J., Muzammil, S., Sweet, R., et al. (2016). Structural diversity in a human antibody germline library. *MAbs* 8(6), 1045-1063. doi: 10.1080/19420862.2016.1190060.

[237] Vargas-Madrazo, E., Lara-Ochoa, F., and Almagro, J.C. (1995). Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. *J Mol Biol* 254(3), 497-504.

[238] Vaughan, T.J., Williams, A.J., Pritchard, K., Osbourn, J.K., Pope, A.R., Earnshaw, J.C., et al. (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized

phage display library. *Nat Biotechnol* 14(3), 309-314. doi: 10.1038/nbt0396-309.

All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

Individual Applicant

Street : 4650 Lisann Street
City : San Diego
State : CA
Country : US
PostalCode : 92117
PhoneNumber : 858 945-4740
FaxNumber :
EmailAddress : pvaladon@abdesignlabs.com
<110> LastName : Valadon
<110> FirstName : Philippe
<110> MiddleInitial :
<110> Suffix :

Individual Applicant

Street : 320 Concord Avenue
City : Cambridge
State : MA
Country : US
PostalCode : 02138
PhoneNumber : 617 710-4487
FaxNumber :
EmailAddress : juan.c.almagro@gmail.com
<110> LastName : ALMAGRO
<110> FirstName : Juan Carlos
<110> MiddleInitial :
<110> Suffix :

Application Project

<120> Title : Highly Functional Antibody Libraries
<130> AppFileReference : PCT
<140> CurrentAppNumber :
<141> CurrentFilingDate : ____-__-__

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gaaattgtgt tgacgcagtc tccaggcacc ctgtctttgt ctccagggga acgtgccacc 60

ctctcctgcc gtgccagtca gagtgttagc agcagctact tagcctggta tcagcagaaa 120

cctggccagg ctccccgact cctcatctat ggtgcatcta gccgtgccac tggtatccca 180

gaccgtttca gtggcagtgg gtctgggaca gacttcactc tcaccatcag cagactggag 240

cctgaagatt ttgcagtgta ttactgtcag cagtatggta gctcacctct gacgttcggc 300

caaggtacca aggtggaaat caaa 324
<212> Type : DNA
<211> Length : 324
SequenceName : 1
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gacatcgtga tgacccagtc tccagactcc ctggctgtgt ctctgggcga gcgtgccacc 60

atcaactgca agtccagcca gagtgtttta tacagctcca acaataagaa ctacttagct 120

tggtatcagc agaaaccagg acagcctcct aagctgctca tttactgggc atctacccgg 180

gaatccgggg tccctgaccg attcagtggc agcgggtctg ggacagattt cactctcacc 240

atcagcagcc tgcaggctga agatgtggca gtttattact gtcagcaata ttatagtact 300

cctctgacgt tcggccaagg taccaaggtg gaaatcaaa 339

<212> Type : DNA
<211> Length : 339
SequenceName : 2
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gaggtgcagc tgttggagtc tgggggaggc ttggtacagc ctgggggggtc cctgcgactc 60

tcctgtgcag cctctggatt cacctttagc agctatgcca tgagctgggt ccgccaggct 120

ccagggaagg ggctggagtg ggtgtcagct attagtggta gtggtggtag cacatactac 180

gcagactccg tgaagggccg gttcaccatc tcccgtgaca attccaagaa cacgctgtat 240

ctgcaaatga acagcctgcg tgccgaggac acggccgtgt attactgtgc gaaatacgac 300

ggtatctacg gtgaactgga cttctggggc caaggaaccc tggtcaccgt ttcctca 357

<212> Type : DNA
<211> Length : 357
SequenceName : 3
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSGYDYAEY FQHWGQGTLV TVSS 24
<212> Type : PRT
<211> Length : 24
SequenceName : 4
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSYGYAEYF QHWGQGTLVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 5
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
TTVTAEYFQH WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21

SequenceName : 6
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YSGSYYAEYF QHWGQGTLVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 7
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYGDYAEYFQ HWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 8
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYSNYAEYFQ HWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 9
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
SIAARAEYFQ HWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 10
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VQLERAEYFQ HWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 11
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYDILTGYYN AEYFQHWGQG TLVTVSS 27
<212> Type : PRT
<211> Length : 27
SequenceName : 12
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYDSSGYYY AEYFQHWGQG TLVTVSS 27
<212> Type : PRT
<211> Length : 27
SequenceName : 13
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYGSGSYYN AEYFQHWGQG TLVTVSS 27
<212> Type : PRT
<211> Length : 27
SequenceName : 14
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
EYSSSSAEYF QHWGQGTLVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 15
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GITGTAEYFQ HWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 16
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GIVGATAEYF QHWGQGTLVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 17
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GTTGTAEYFQ HWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 18
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence

<400> PreSequenceString :
GYSSGYAEYF QHWGQGTLVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 19
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
LTGAEYFQHW GQGTLVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 20
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VDIVATIAEY FQHWGQGTLV TVSS 24
<212> Type : PRT
<211> Length : 24
SequenceName : 21
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSGYDYDAF DVWGQGTMVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 22
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSYGYDAFD VWGQGTMVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 23
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
TTVTDAFDVW GQGTMVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 24
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YSGSYYDAFD VWGQGTMVTV SS 22
<212> Type : PRT

<211> Length : 22
SequenceName : 25
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYGDYDAFDV WGQGTMVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 26
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYSNYDAFDV WGQGTMVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 27
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
SIAARDAFDV WGQGTMVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 28
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VQLERDAFDV WGQGTMVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 29
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYDILTGYYN DAFDVWGQGT MVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 30
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYDSSGYYY DAFDVWGQGT MVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 31
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYGSGSYYN DAFDVWGQGT MVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 32
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
EYSSSSDAFD VWGQGTMVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 33
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GITGTDAFDV WGQGTMVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 34
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GIVGATDAFD VWGQGTMVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 35
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GTTGTDAFDV WGQGTMVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 36
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSSGYDAFD VWGQGTMVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 37
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
LTGDAFDVWG QGTMVTVSS 19
<212> Type : PRT
<211> Length : 19
SequenceName : 38
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VDIVATIDAF DVWGQGTMVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 39
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSGYDYYFD YWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 40
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSYGYYFDY WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 41
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
TTVTYFDYWG QGTLVTVSS 19
<212> Type : PRT
<211> Length : 19
SequenceName : 42
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YSGSYYYFDY WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 43
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYGDYYFDYW GQGTLVTVSS 20

<212> Type : PRT
<211> Length : 20
SequenceName : 44
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYSNYYFDYW GQGTLVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 45
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
SIAARYFDYW GQGTLVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 46
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VQLERYFDYW GQGTLVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 47
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYDILTGYYN YFDYWGQGTL VTVSS 25
<212> Type : PRT
<211> Length : 25
SequenceName : 48
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYDSSGYYY YFDYWGQGTL VTVSS 25
<212> Type : PRT
<211> Length : 25
SequenceName : 49
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYGSGSYYN YFDYWGQGTL VTVSS 25
<212> Type : PRT
<211> Length : 25
SequenceName : 50

SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
EYSSSSYFDY WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 51
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GITGTYFDYW GQGTLVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 52
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GIVGATYFDY WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 53
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GTTGTYFDYW GQGTLVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 54
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSSGYYFDY WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 55
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
LTGYFDYWGQ GTLVTVSS 18
<212> Type : PRT
<211> Length : 18
SequenceName : 56
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VDIVATIYFD YWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 57
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSGYDYNWF DSWGQGTLVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 58
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSYGYNWFD SWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 59
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
TTVTNWFDSW GQGTLVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 60
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YSGSYYNWFD SWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 61
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYGDYNWFDS WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 62
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :

DYSNYNWFDS WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 63
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
SIAARNWFDS WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 64
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VQLERNWFDS WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 65
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYDILTGYYN NWFDSWGQGT LVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 66
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYDSSGYYY NWFDSWGQGT LVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 67
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYGSGSYYN NWFDSWGQGT LVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 68
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
EYSSSSNWFD SWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22

SequenceName : 69
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GITGTNWFDS WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 70
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GIVGATNWFD SWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 71
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GTTGTNWFDS WGQGTLVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 72
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSSGYNWFD SWGQGTLVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 73
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
LTGNWFDSWG QGTLVTVSS 19
<212> Type : PRT
<211> Length : 19
SequenceName : 74
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VDIVATINWF DSWGQGTLVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 75
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSGYDYYGM DVWGQGTTVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 76
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSYGYYGMD VWGQGTTVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 77
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
TTVTYGMDVW GQGTTVTVSS 20
<212> Type : PRT
<211> Length : 20
SequenceName : 78
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YSGSYYYGMD VWGQGTTVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 79
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYGDYYGMDV WGQGTTVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 80
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
DYSNYYGMDV WGQGTTVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 81
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence

<400> PreSequenceString :
SIAARYGMDV WGQGTTVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 82
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VQLERYGMDV WGQGTTVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 83
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYDILTGYYN YGMDVWGQGT TVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 84
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYDSSGYYY YGMDVWGQGT TVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 85
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
YYYGSGSYYN YGMDVWGQGT TVTVSS 26
<212> Type : PRT
<211> Length : 26
SequenceName : 86
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
EYSSSSYGMD VWGQGTTVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 87
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GITGTYGMDV WGQGTTVTVS S 21
<212> Type : PRT

<211> Length : 21
SequenceName : 88
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GIVGATYGMD VWGQGTTVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 89
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GTTGTYGMDV WGQGTTVTVS S 21
<212> Type : PRT
<211> Length : 21
SequenceName : 90
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
GYSSGYYGMD VWGQGTTVTV SS 22
<212> Type : PRT
<211> Length : 22
SequenceName : 91
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
LTGYGMDVWG QGTTVTVSS 19
<212> Type : PRT
<211> Length : 19
SequenceName : 92
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
VDIVATIYGM DVWGQGTTVT VSS 23
<212> Type : PRT
<211> Length : 23
SequenceName : 93
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gacacggcyg tgtattactg tgc 23
<212> Type : DNA
<211> Length : 23
SequenceName : 94
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
tgttggcctc ccgggcctga agagacggtg accattgtcc 40
<212> Type : DNA
<211> Length : 40
SequenceName : 95
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
tgttggcctc ccgggcctga ggagacggtg accgtggtc 39
<212> Type : DNA
<211> Length : 39
SequenceName : 96
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
tgttggcctc ccgggcctga ggagacrgtg accagggt 38
<212> Type : DNA
<211> Length : 38
SequenceName : 97
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
cgctggattg ttattactcg cg 22
<212> Type : DNA
<211> Length : 22
SequenceName : 98
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
cacaggtctc gtgtcctcgg cacgcaggct gttcatttg 39
<212> Type : DNA
<211> Length : 39
SequenceName : 99
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gtgtggtctc ggacacggcy gtgtattact gtgc 34
<212> Type : DNA
<211> Length : 34
SequenceName : 100
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
tcgcggccca gccggccatg 20
<212> Type : DNA
<211> Length : 20
SequenceName : 101
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gtgttggcct cccgggcctg 20
<212> Type : DNA
<211> Length : 20
SequenceName : 102
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
caggtgcagc tggtgcagtc tggggctgag gtgaagaagc ctggggcctc agtgaaggtc 60

tcctgcaagg cttctggata caccttcacc ggctactata tgcactgggt gcgacaggcc 120

cctggacaag ggcttgagtg gatgggacgg atcaacccta acagtggtgg cacaaactat 180

gcacagaagt ttcagggcag ggtcaccagt accagggaca cgtccatcag cacagcctac 240

atggagctga gcaggctgag atctgacgac acggtcgtgt attactgtgc gagaga 296

<212> Type : DNA
<211> Length : 296
SequenceName : 103
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
caggttcagc tggtgcagtc tggagctgag gtgaagaagc ctggggcctc agtgaaggtc 60

tcctgcaagg cttctggtta cacctttacc agctatggta tcagctgggt gcgacaggcc 120

cctggacaag ggcttgagtg gatgggatgg atcagcgctt acaatggtaa cacaaactat 180

gcacagaagc tccagggcag agtcaccatg accacagaca catccacgag cacagcctac 240

atggagctga ggagcctgag atctgacgac acggccgtgt attactgtgc gagaga 296

<212> Type : DNA
<211> Length : 296
SequenceName : 104
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :

caggtgcagc tggtgcagtc tggggctgag gtgaagaagc ctggggcctc agtgaaggtt 60

tcctgcaagg catctggata caccttcacc agctactata tgcactgggt gcgacaggcc 120

cctggacaag ggcttgagtg gatgggaata atcaacccta gtggtggtag cacaagctac 180

gcacagaagt tccagggcag agtcaccatg accagggaca cgtccacgag cacagtctac 240

atggagctga gcagcctgag atctgaggac acggccgtgt attactgtgc gagaga 296

<212> Type : DNA
<211> Length : 296
SequenceName : 105
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
caggtgcagc tggtgcagtc tggggctgag gtgaagaagc ctgggtcctc ggtgaaggtc 60

tcctgcaagg cttctggagg caccttcagc agctatgcta tcagctgggt gcgacaggcc 120

cctggacaag ggcttgagtg gatgggaggg atcatcccta tctttggtac agcaaactac 180

gcacagaagt tccagggcag agtcacgatt accgcggacg aatccacgag cacagcctac 240

atggagctga gcagcctgag atctgaggac acggccgtgt attactgtgc gagaga 296

<212> Type : DNA
<211> Length : 296
SequenceName : 106
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gaggtgcagc tgttggagtc tgggggaggc ttggtacagc ctgggggggtc cctgagactc 60

tcctgtgcag cctctggatt cacctttagc agctatgcca tgagctgggt ccgccaggct 120

ccagggaagg ggctggagtg ggtctcagct attagtggta gtggtggtag cacatactac 180

gcagactccg tgaagggccg gttcaccatc tccagagaca attccaagaa cacgctgtat 240

ctgcaaatga acagcctgag agccgaggac acggccgtat attactgtgc gaaaga 296

<212> Type : DNA
<211> Length : 296
SequenceName : 107
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
caggtgcagc tggtggagtc tgggggaggc gtggtccagc ctgggaggtc cctgagactc 60

tcctgtgcag cctctggatt caccttcagt agctatgcta tgcactgggt ccgccaggct 120

ccaggcaagg ggctagagtg ggtggcagtt atatcatatg atggaagtaa taaatactac 180

gcagactccg tgaagggccg attcaccatc tccagagaca attccaagaa cacgctgtat 240

ctgcaaatga acagcctgag agctgaggac acggctgtgt attactgtgc gagaga 296

<212> Type : DNA
<211> Length : 296
SequenceName : 108
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gaggtgcagc tggtggagtc tgggggaggc ttggtacagc ctgggggtc cctgagactc 60

tcctgtgcag cctctggatt caccttcagt agctatagca tgaactgggt ccgccaggct 120

ccagggaagg ggctggagtg ggtttcatac attagtagta gtagtagtac catatactac 180

gcagactctg tgaagggccg attcaccatc tccagagaca atgccaagaa ctcactgtat 240

ctgcaaatga acagcctgag agccgaggac acggctgtgt attactgtgc gagaga 296

<212> Type : DNA
<211> Length : 296
SequenceName : 109
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
caggtgcagc tacagcagtg gggcgcagga ctgttgaagc cttcggagac cctgtccctc 60

acctgcgctg tctatggtgg gtccttcagt ggttactact ggagctggat ccgccagccc 120

ccagggaagg ggctggagtg gattggggaa atcaatcata gtggaagcac caactacaac 180

ccgtccctca agagtcgagt caccatatca gtagacacgt ccaagaacca gttctccctg 240

aagctgagct ctgtgaccgc cgcggacacg gctgtgtatt actgtgcgag agg 293

<212> Type : DNA
<211> Length : 293
SequenceName : 110
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
caggtgcagc tgcaggagtc gggcccagga ctggtgaagc cttcggagac cctgtccctc 60

acctgcactg tctctggtgg ctccatcagt agttactact ggagctggat ccggcagccc 120

ccagggaagg gactggagtg gattgggtat atctattaca gtgggagcac caactacaac 180

ccctccctca agagtcgagt caccatatca gtagacacgt ccaagaacca gttctccctg 240

aagctgagct ctgtgaccgc tgcggacacg gccgtgtatt actgtgcgag aga 293

<212> Type : DNA
<211> Length : 293
SequenceName : 111
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gaggtgcagc tggtgcagtc tggagcagag gtgaaaaagc ccggggagtc tctgaagatc 60

tcctgtaagg gttctggata cagctttacc agctactgga tcggctgggt gcgccagatg 120

cccgggaaag gcctggagtg gatggggatc atctatcctg gtgactctga taccagatac 180

agcccgtcct tccaaggcca ggtcaccatc tcagccgaca agtccatcag caccgcctac 240

ctgcagtgga gcagcctgaa ggcctcggac accgccatgt attactgtgc gagaca 296

<212> Type : DNA
<211> Length : 296
SequenceName : 112
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
caggtacagc tgcagcagtc aggtccagga ctggtgaagc cctcgcagac cctctcactc 60

acctgtgcca tctccgggga cagtgtctct agcaacagtg ctgcttggaa ctggatcagg 120

cagtccccat cgagaggcct tgagtggctg ggaaggacat actacaggtc caagtggtat 180

aatgattatg cagtatctgt gaaaagtcga ataaccatca acccagacac atccaagaac 240

cagttctccc tgcagctgaa ctctgtgact cccgaggaca cggctgtgta ttactgtgca 300

agaga 305

<212> Type : DNA
<211> Length : 305
SequenceName : 113
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gacatccaga tgacccagtc tccttccacc ctgtctgcat ctgtaggaga cagagtcacc 60

atcacttgcc gggccagtca gagtattagt agctggttgg cctggtatca gcagaaacca 120

gggaaagccc ctaagctcct gatctatgat gcctccagtt tggaaagtgg ggtcccatca 180

aggttcagcg gcagtggatc tgggacagaa ttcactctca ccatcagcag cctgcagcct 240

gatgattttg caacttatta ctgccaacag tataatagtt attctcc 287
<212> Type : DNA
<211> Length : 287
SequenceName : 114
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gacatccaga tgacccagtc tccatcttcc gtgtctgcat ctgtaggaga cagagtcacc 60

atcacttgtc gggcgagtca gggtattagc agctggttag cctggtatca gcagaaacca 120

gggaaagccc ctaagctcct gatctatgct gcatccagtt tgcaaagtgg ggtcccatca 180

aggttcagcg gcagtggatc tgggacagat ttcactctca ccatcagcag cctgcagcct 240

gaagattttg caacttacta ttgtcaacag gctaacagtt tccctcc 287
<212> Type : DNA
<211> Length : 287
SequenceName : 115
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gacatccaga tgacccagtc tccatcctcc ctgtctgcat ctgtaggaga cagagtcacc 60

atcacttgcc aggcgagtca ggacattagc aactatttaa attggtatca gcagaaacca 120

gggaaagccc ctaagctcct gatctacgat gcatccaatt tggaaacagg ggtcccatca 180

aggttcagtg gaagtggatc tgggacagat tttactttca ccatcagcag cctgcagcct 240

gaagatattg caacatatta ctgtcaacag tatgataatc tccctcc 287
<212> Type : DNA
<211> Length : 287
SequenceName : 116
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gatattgtga tgactcagtc tccactctcc ctgcccgtca cccctggaga gccggcctcc 60

atctcctgca ggtctagtca gagcctcctg catagtaatg gatacaacta tttggattgg 120

tacctgcaga agccagggca gtctccacag ctcctgatct atttgggttc taatcgggcc 180

tccggggtcc ctgacaggtt cagtggcagt ggatcaggca cagattttac actgaaaatc 240

agcagagtgg aggctgagga tgttggggtt tattactgca tgcaagctct acaaactcct 300

cc 302
<212> Type : DNA
<211> Length : 302
SequenceName : 117
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gacatccaga tgacccagtc tccatcctcc ctgtctgcat ctgtaggaga cagagtcacc 60

atcacttgcc gggcaagtca gagcattagc agctatttaa attggtatca gcagaaacca 120

gggaaagccc ctaagctcct gatctatgct gcatccagtt tgcaaagtgg ggtcccatca 180

aggttcagtg gcagtggatc tgggacagat ttcactctca ccatcagcag tctgcaacct 240

gaagattttg caacttacta ctgtcaacag agttacagta cccctcc 287
<212> Type : DNA
<211> Length : 287
SequenceName : 118
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gaaattgtgt tgacacagtc tccagccacc ctgtctttgt ctccagggga aagagccacc 60

ctctcctgca gggccagtca gagtgttagc agctacttag cctggtacca acagaaacct 120

ggccaggctc ccaggctcct catctatgat gcatccaaca gggccactgg catcccagcc 180

aggttcagtg gcagtgggtc tgggacagac ttcactctca ccatcagcag cctagagcct 240

gaagattttg cagtttatta ctgtcagcag cgtagcaact ggcctcc 287
<212> Type : DNA
<211> Length : 287
SequenceName : 119
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gaaatagtga tgacgcagtc tccagccacc ctgtctgtgt ctccagggga aagagccacc 60

ctctcctgca gggccagtca gagtgttagc agcaacttag cctggtacca gcagaaacct 120

ggccaggctc ccaggctcct catctatggt gcatccacca gggccactgg tatcccagcc 180

aggttcagtg gcagtgggtc tgggacagag ttcactctca ccatcagcag cctgcagtct 240

gaagattttg cagtttatta ctgtcagcag tataataact ggcctcc 287
<212> Type : DNA
<211> Length : 287
SequenceName : 120
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gaaattgtgt tgacgcagtc tccaggcacc ctgtctttgt ctccagggga aagagccacc 60

ctctcctgca gggccagtca gagtgttagc agcagctact tagcctggta ccagcagaaa 120

cctggccagg ctcccaggct cctcatctat ggtgcatcca gcagggccac tggcatccca 180

gacaggttca gtggcagtgg gtctgggaca gacttcactc tcaccatcag cagactggag 240

cctgaagatt ttgcagtgta ttactgtcag cagtatggta gctcacctcc 290

<212> Type : DNA
<211> Length : 290
SequenceName : 121
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
gacatcgtga tgacccagtc tccagactcc ctggctgtgt ctctgggcga gagggccacc 60

atcaactgca agtccagcca gagtgtttta tacagctcca acaataagaa ctacttagct 120

tggtaccagc agaaaccagg acagcctcct aagctgctca tttactgggc atctacccgg 180

gaatccgggg tccctgaccg attcagtggc agcgggtctg ggacagattt cactctcacc 240

atcagcagcc tgcaggctga agatgtggca gtttattact gtcagcaata ttatagtact 300

cctcc 305

<212> Type : DNA
<211> Length : 305
SequenceName : 122
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
tcttctgagc tgactcagga ccctgctgtg tctgtggcct tgggacagac agtcaggatc 60
acatgccaag gagacagcct cagaagctat tatgcaagct ggtaccagca gaagccagga 120
caggcccctg tacttgtcat ctatggtaaa aacaaccggc cctcagggat cccagaccga 180
ttctctggct ccagctcagg aaacacagct tccttgacca tcactggggc tcaggcggaa 240
gatgaggctg actattactg taactcccgg gacagcagtg gtaaccatct 290

<212> Type : DNA
<211> Length : 290
SequenceName : 123
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
aattttatgc tgactcagcc ccactctgtg tcggagtctc cggggaagac ggtaaccatc 60

tcctgcaccc gcagcagtgg cagcattgcc agcaactatg tgcagtggta ccagcagcgc 120

ccgggcagtt cccccaccac tgtgatctat gaggataacc aaagaccctc tggggtccct 180

gatcggttct ctggctccat cgacagctcc tccaactctg cctccctcac catctctgga 240

ctgaagactg aggacgaggc tgactactac tgtcagtctt atgatagcag caatca 296

<212> Type : DNA
<211> Length : 296
SequenceName : 124
SequenceDescription :

Sequence

<213> OrganismName : Homo sapiens
<400> PreSequenceString :
tcctatgtgc tgactcagcc accctcagtg tcagtggccc caggaaagac ggccaggatt 60

acctgtgggg gaaacaacat tggaagtaaa agtgtgcact ggtaccagca gaagccaggc 120

caggcccctg tgctggtcat ctattatgat agcgaccggc cctcagggat ccctgagcga 180

ttctctggct ccaactctgg gaacacggcc accctgacca tcagcagggt cgaagccggg 240

gatgaggccg actattactg tcaggtgtgg gacagtagta gtgatcatcc 290
<212> Type : DNA
<211> Length : 290
SequenceName : 125
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
tctttcccta cacgacgctc ttccgatctt ggcagaaatt gtgttgacgc ag 52

<212> Type : DNA
<211> Length : 52
SequenceName : 126
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
tctttcccta cacgacgctc ttccgatctg actccctggc tgtgtctct 49
<212> Type : DNA
<211> Length : 49
SequenceName : 127
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gtgactggag ttcagacgtg tgctcttccg atcccttggt accttggccg aac 53

<212> Type : DNA
<211> Length : 53
SequenceName : 128
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
tctttcccta cacgacgctc ttccgatctg tgcagcctct ggattcacc 49
<212> Type : DNA
<211> Length : 49
SequenceName : 129
SequenceDescription :

Sequence

<213> OrganismName : artificial sequence
<400> PreSequenceString :
gtgactggag ttcagacgtg tgctcttccg atcgtggtga tggtgttggc ctc 53

<212> Type : DNA
<211> Length : 53
SequenceName : 130
SequenceDescription :

CLAIMS

What is claimed is:

1. A method of preparing a highly functional library of antibody-encoding polynucleotides, which comprises the steps of:
(a) preparing a primary library of nucleic acid sequences that express and display antibody-encoding polynucleotides, wherein each nucleic acid sequence encodes an antibody containing a heavy chain variable region;
(b) displaying translated products of nucleic acid sequences as antibody fragments and applying diverse conditions for selecting antibody fragments of improved developability;
(c) amplifying antibody-encoding polynucleotides from a pool of antibodies selected in step (b); and
(d) preparing a secondary library of antibody-encoding polynucleotides by replacing CDR-H3 of each heavy chain variable region of antibody-encoding polynucleotides amplified in step (c) by a CDR-H3 having a different sequence from that which it replaces.

2. The method of claim 1, wherein a plurality of nucleic acid sequences in the primary library encodes at least 10^3 different antibody-encoding polynucleotides.

3. The method of claim 1, wherein a plurality of nucleic acid sequences in the primary library encode at least 10^4 different antibody-encoding polynucleotides.

4. The method of claim 1, wherein a plurality of nucleic acid sequences in the primary library encode at least 10^5 different antibody-encoding polynucleotides.

5. The method of claim 1, wherein the primary library encodes for an antibody fragment selected from Fab, scFv, V_HH or single V_H domain.

6. The method of claim 1, wherein the primary library encodes for a subset of human scaffolds.

7. The method of claim 1, wherein the primary library encodes for a subset of human scaffolds selected from VH1-02 (SEQA ID NO: 103); VH1-18 (SEQA ID NO: 104); VH1-46 (SEQA ID NO: 105); VH1-69 (SEQA ID NO: 106); VH3-23 (SEQA ID NO: 107); VH3-30/33 (SEQA ID NO: 108); VH3-48 (SEQA ID NO: 109); VH4-34 (SEQA ID NO: 110); VH4-59/61 (SEQA ID NO: 111); VH5-51 (SEQA ID NO: 112); VH6-1 (SEQA ID NO: 113); VK1-5 (SEQA ID NO: 114); VK1-12 (SEQA ID NO: 115); VK1-33 (SEQA ID NO: 116); VK2-28 (SEQA ID NO: 117); VK1-39 (SEQA ID NO: 118); VK3-11 (SEQA ID NO: 119); VK3-15 (SEQA ID NO: 120); VK3-20 (SEQA ID NO: 121); VK4-1 (SEQA ID NO: 122); VL3-19 (SEQA ID NO: 123); VL6-57 (SEQA ID NO: 124); VL3-21 (SEQA ID NO: 125).

8. The method of claim 1, wherein the heavy chains of the primary library are encoded by a subset of human heavy chain scaffolds from human IGHV-3 germline gene family.

9. The method of claim 1, wherein the display method is selected from bacteriophage display, RNA display, yeast display, bacterial cell surface display, and mammalian cell surface display.

10. The method of claim 1, wherein the display method is M13 phage display.

11. The method of claim 1, wherein selection of antibody fragments of increased developability is based on affinity to a folded antibody fragment.

12. The method of claim 1, wherein selection of folded antibody fragments of increased developability is based on binding of V_H to Protein A.

13. The method of claim 1, wherein selection of folded antibody fragments of increased developability is based on binding of members of human VK2 or VK4 families to Protein-L.

14. The method of claim 1, wherein selection of antibody fragments of increased developability is based on protein unfolding using treatments selected from high temperature, acid treatment, basic treatment, protease sensitivity, stability in serum, denaturation reagents, urea, and high salt concentration.

15. The method of claim 1, wherein selection of antibody fragments of increased developability is based on unfolding at high temperature.

16. The method of claim 1, wherein selection of antibody fragments of increased developability is based on removal of antibodies of decreased developability using methods selected from non-specific absorption, binding to HSP90, and hydrophobic interactions.

17. The method of claim 1, wherein CDR-H3 sequences of the primary library are selected from a set of IGDH germline genes.

18. The method of claim 1, wherein CDR-H3 of the primary library is selected from a set of IGDH germline genes with translations taken from GYSGYDY, GYSYGY, TTVT, YSGSYY, DYGDY, DYSNY, SIAAR, VQLER, YYDILTGYYN, YYYDSSGYYY, YYYGSGSYYN, EYSSSS, GITGT, GIVGAT, GTTGT, GYSSGY, LTG, VDIVATI.

19. The method of claim 1, wherein the CDR-H3 sequence of the heavy chain of the primary library is a unique CDR-H3 sequence.

20. The method of claim 1, wherein the CDR-H3 of the primary library is replaced with human natural CDR-H3.

21. The method of claim 1, wherein the heavy chain CDR-H3/JH fragment is replaced in the pool of selected antibodies.

22. The method of claim 1, wherein the CDR-H3 of the primary library is replaced with synthetic CDR-H3.

FIGURE 1

FIGURE 1A:

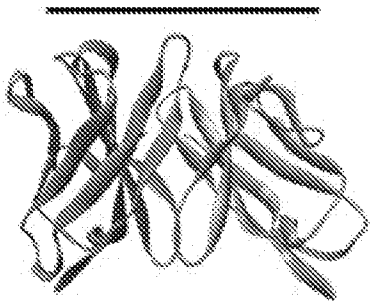

FIGURE 1B:

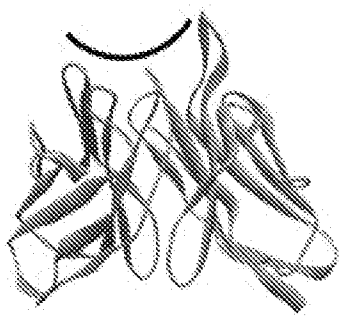

FIGURE 2

1	2vxt_H	31	3stz_A	61	4j6r_H	91	4uta_H
2	2w9e_H	32	3u30_C	62	4jqi_H	92	4xak_D
3	2xqy_G	33	3u7y_H	63	4jr9_H	93	4xmp_H
4	2xra_H	34	3wfd_H	64	4jzj_A	94	4xnu_H
5	2xtj_D	35	3wih_H	65	4k2u_H	95	4xnx_H
6	2yc1_A	36	3x3f_H	66	4k3j_H	96	4xtr_C
7	2ypv_H	37	3ztn_H	67	4k94_H	97	4xvt_H
8	2zch_H	38	4ag4_H	68	4ki5_C	98	4zfg_H
9	3bgf_H	39	4al8_H	69	4kuc_F	99	4zfo_H
10	3bt2_H	40	4bz1_H	70	4kv5_E	100	5czx_H
11	3d85_B	41	4bz2_H	71	4l5f_H	101	5d8j_H
12	3dvg_B	42	4d9q_H	72	4leo_A	102	5dfv_C
13	3ehb_C	43	4dke_H	73	4lqf_H	103	5dfw_H
14	3ejz_C	44	4dkf_H	74	4lu5_H	104	5dum_H
15	3iu3_A	45	4dn4_H	75	4lvh_B	105	5e8d_H
16	3kr3_H	46	4dtg_H	76	4m8q_A	106	5e8e_B
17	3l95_H	47	4dvr_H	77	4mwf_A	107	5en2_A
18	3lhp_H	48	4f15_B	78	4n9g_A	108	5f96_H
19	3liz_H	49	4f2m_A	79	4nnp_H	109	5fb8_B
20	3mxw_H	50	4f3f_B	80	4np4_H	110	5ies_H
21	3nfp_A	51	4ffv_H	81	4nx3_B	111	5ikc_B
22	3p0y_H	52	4ffy_H	82	4ogy_H	112	5jhl_H
23	3pgf_H	53	4g6j_H	83	4oii_H	113	5kn5_A
24	3pnw_B	54	4g7v_H	84	4okv_A	114	5kvd_H
25	3pp4_H	55	4gms_H	85	4plj_H	115	5kve_H
26	3r1g_H	56	4h88_H	86	4py8_I	116	5kvf_H
27	3sdy_H	57	4hkx_A	87	4qti_H	117	5kvg_H
28	3skj_H	58	4hkz_B	88	4r8w_H		
29	3sob_H	59	4hwb_H	89	4rdq_G		
30	3sqo_H	60	4i9w_E	90	4u0r_B		

FIGURE 3

#	wt	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y		Diversity
30	S																50	50				100	2
31	S			30									20				50					100	3
32	Y			10													10	20			60	100	4
33	A	40					10								10					10	30	100	5
50	A	10		10	20		10		5							20			5	10	10	100	9
51	I																						
52	S			25											15		35				25	100	4
52a	G																						
53	S			5	10		20									10	30				25	100	6
54	G																						
55	G			15			60										15				10	100	4
56	S			15			15								10		30	10			20	100	6
57	T																						
58	Y			10						10			30			10	10				30	100	6

FIGURE 4

#	wt	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y		Diversity
30	S			25											25		50					100	3
30a	S	25						25									50					100	3
31	S												30				50	20				100	3
32	Y	20											10				20				50	100	4
50	G	20		15			50													15		100	4
91	Y	10		10		5	10	5								10	10				40	100	8
92	G				10		40									10	30				10	100	5
93	S				10								10		10	10	50				10	100	6
94	S	10			5								10			5	50			10	10	100	7
95	P																						
96	L					20			20		20									20	20	100	5

FIGURE 5

#	wt	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y		Diveristy
30a	Y			10			10	5								10	15				50	100	6
30b	S																						
30c	S			10			20						10				40				20	100	5
30d	N																						
30e	N																						
30f	K				40					50											10	100	3
31	N																						
32	Y			15									15			20					50	100	4
50	W	20		15			15													50		100	4
91	Y						10	10							10		10			10	50	100	6
92	Y			20											10		10	10			50	100	5
93	S				10								25			10	45				10	100	5
94	T	10			10								10			10	20	30			10	100	7
95	P																						
96	L					20			20		20									20	20	100	5

FIGURE 6

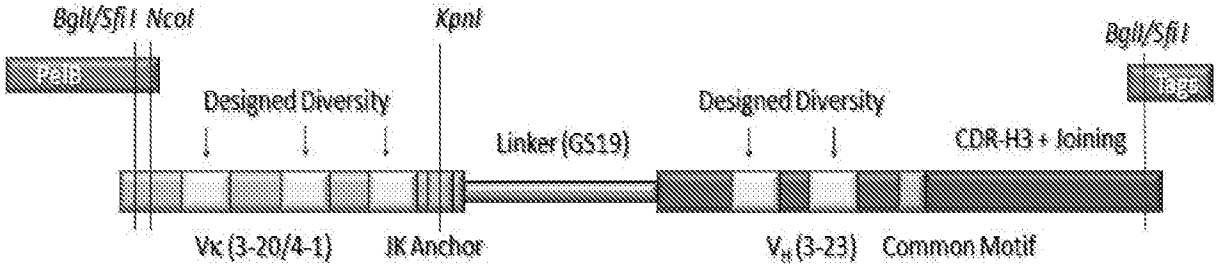
BglI/SfiI NcoI
KpnI
BglI/SfiI
Designed Diversity
Designed Diversity
Linker (GS19)
CDR-H3 + Joining
Vκ (3-20/4-1)
JK Anchor
V_H (3-23)
Common Motif

FIGURE 7

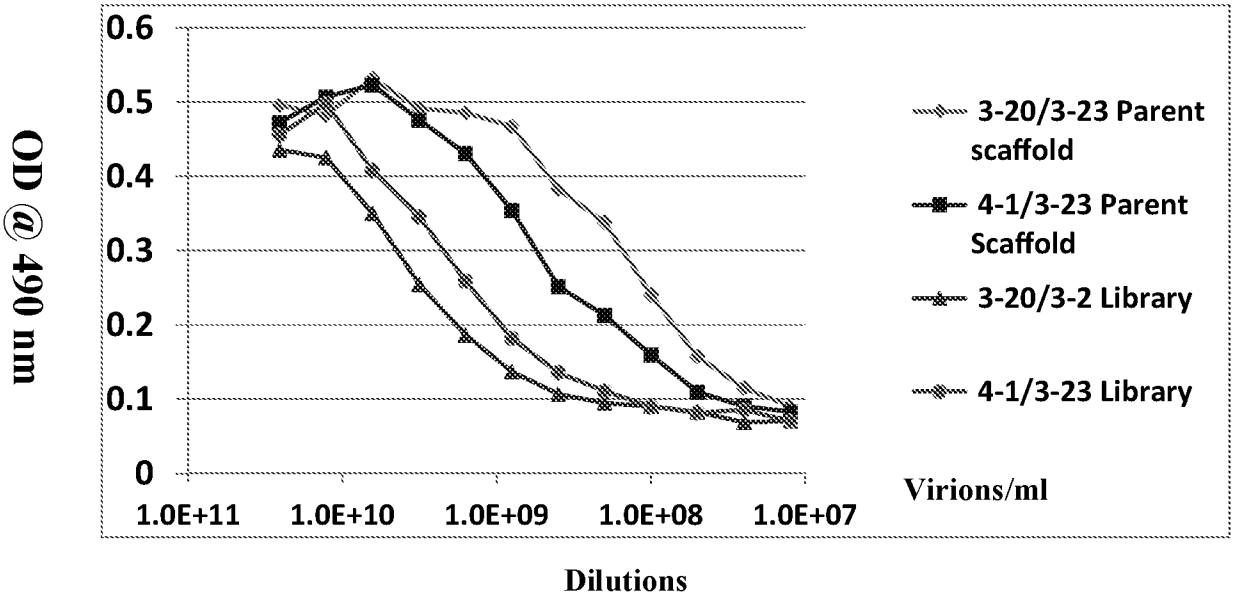
OD @ 490 nm
0.6
0.5
0.4
0.3
0.2
0.1
0
1.0E+11
1.0E+10
1.0E+09
1.0E+08
1.0E+07
Virions/ml
Dilutions
3-20/3-23 Parent scaffold
4-1/3-23 Parent Scaffold
3-20/3-2 Library
4-1/3-23 Library

FIGURE 8

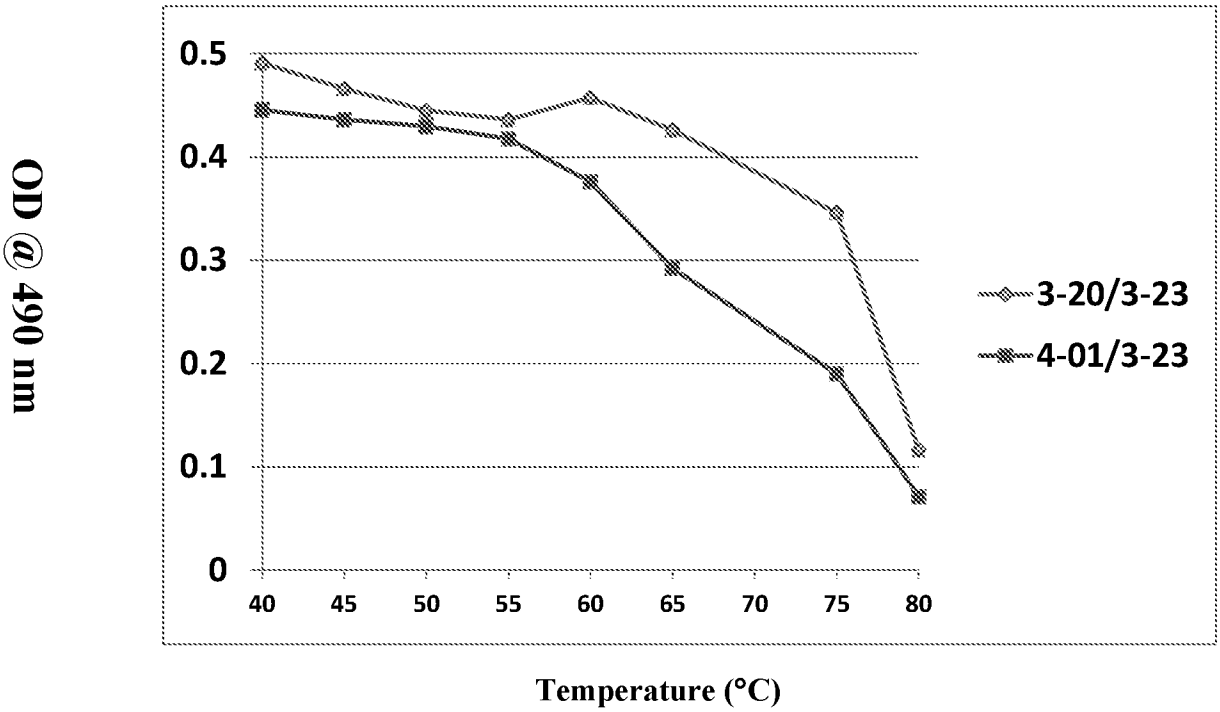
OD @ 490 nm
0.5
0.4
0.3
0.2
0.1
0
40
45
50
55
60
65
70
75
80
3-20/3-23
4-01/3-23
Temperature (°C)

FIGURE 9

FIGURE 9A:

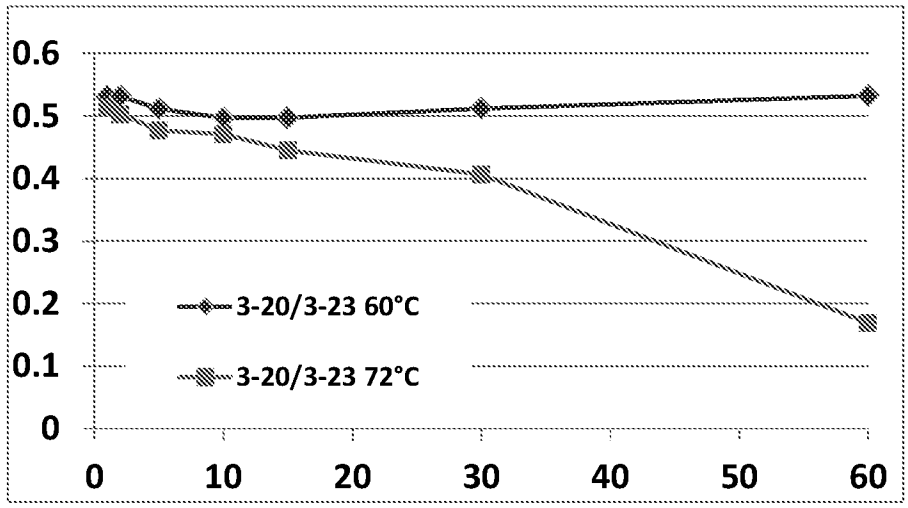

FIGURE 9B:

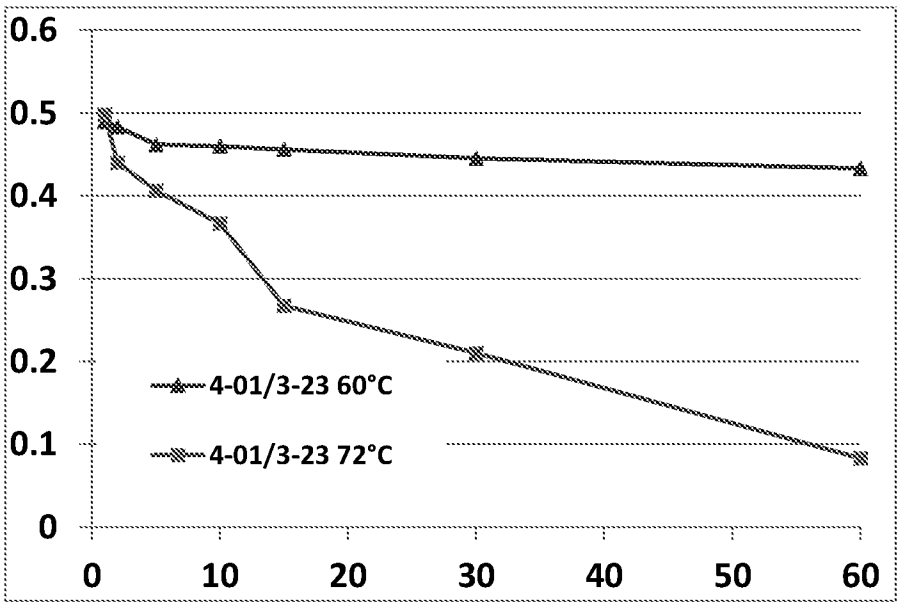

FIGURE 10

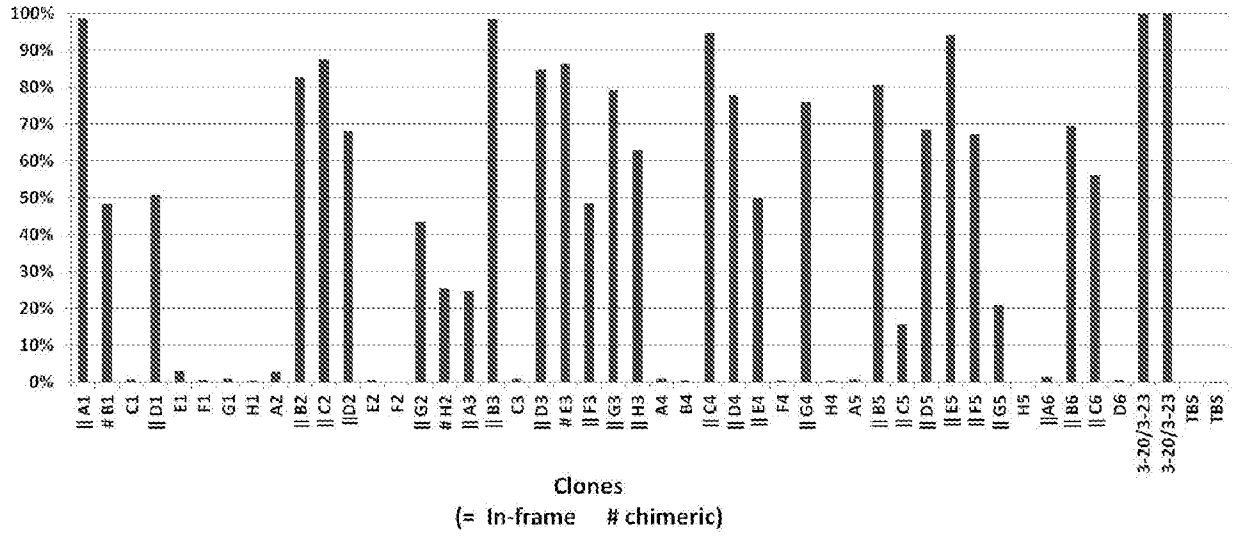
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
B1
C1
D1
E1
F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
H2
A3
B3
C3
D3
E3
F3
G3
H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
3-20/3-23
3-20/3-23
YBS
YBS
Clones
(= In-frame # chimeric)

FIGURE 11

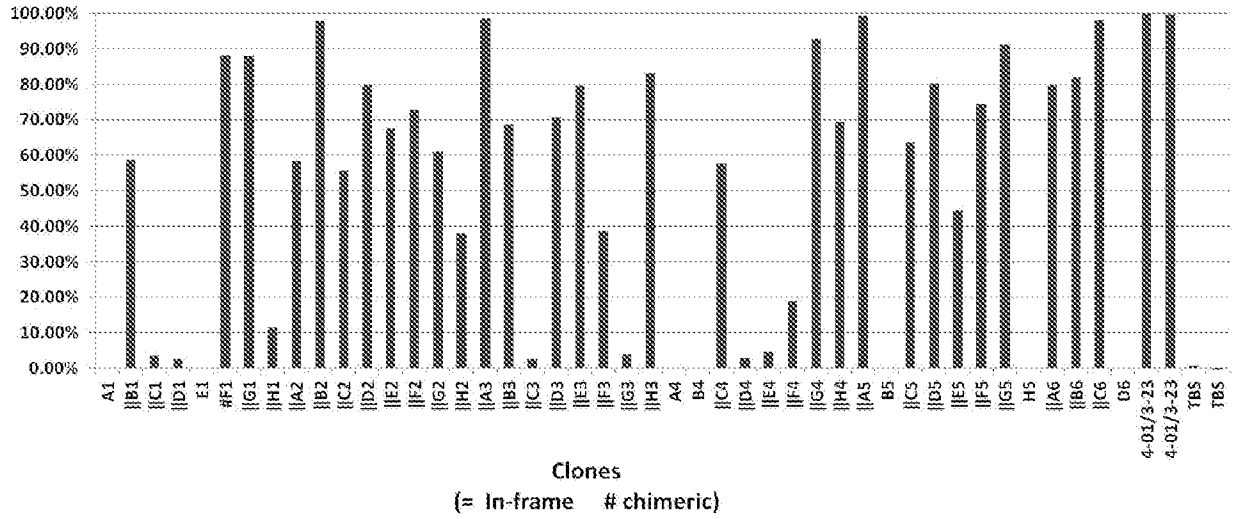
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
A1
B1
C1
D1
E1
F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
H2
A3
B3
C3
D3
E3
F3
G3
H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
4-01/3-23
4-01/3-23
TBS
TBS
Clones
(= In-frame # chimeric)

FIGURE 12

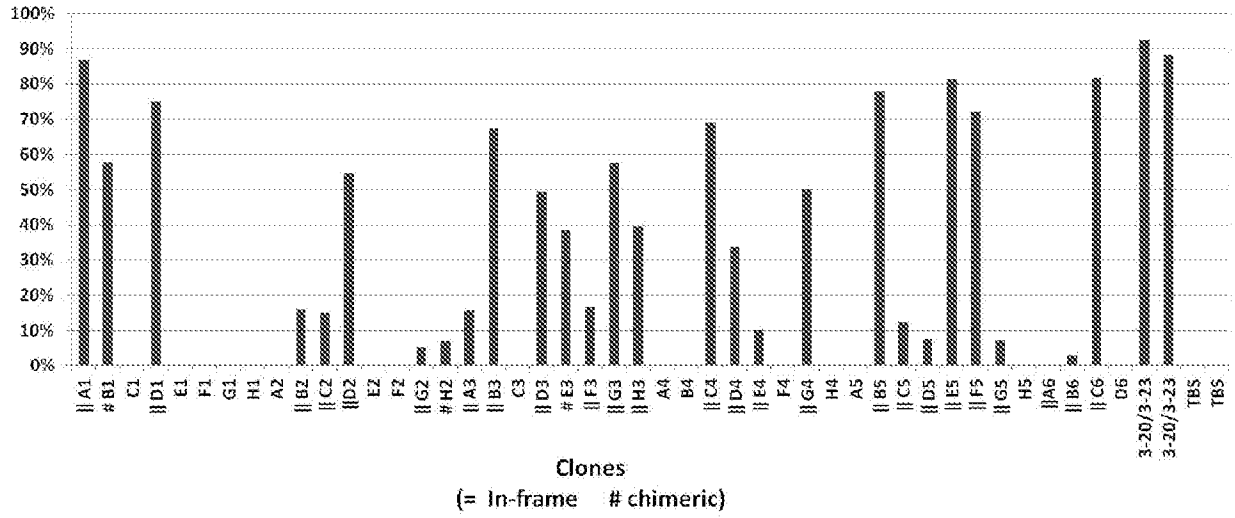
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
B1
C1
D1
E1
F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
H2
A3
B3
C3
D3
E3
F3
G3
H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
3-20/3-23
3-20/3-23
TBS
TBS
Clones
(= In-frame # chimeric)

FIGURE 13

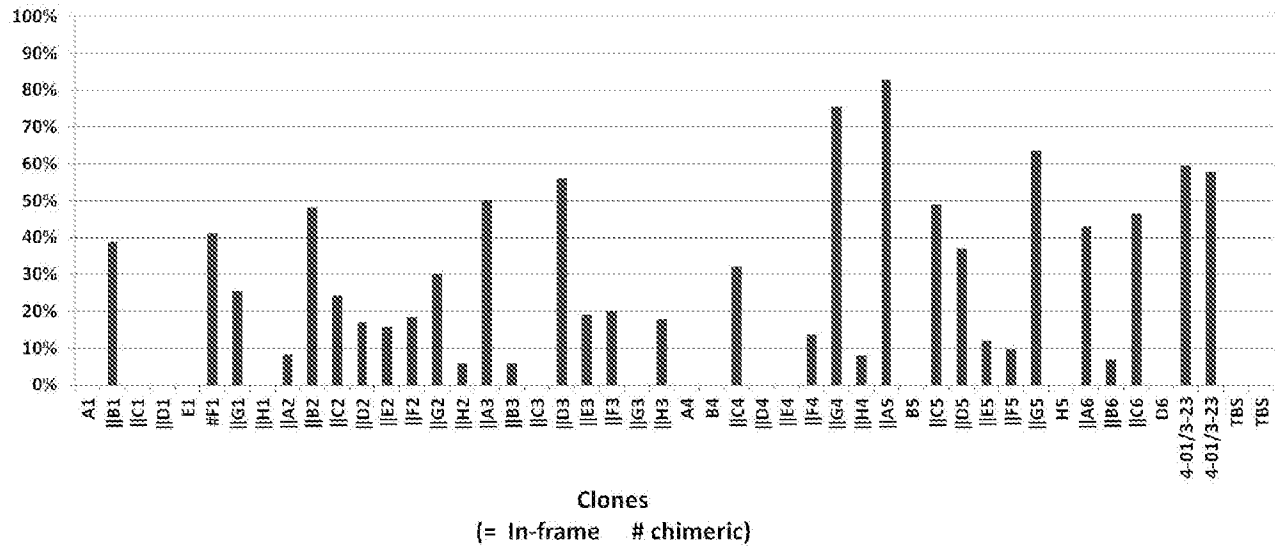
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
B1
C1
D1
E1
F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
H2
A3
B3
C3
D3
E3
F3
G3
H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
4-01/3-23
4-01/3-23
TBS
TBS
Clones
(= In-frame # chimeric)

FIGURE 14

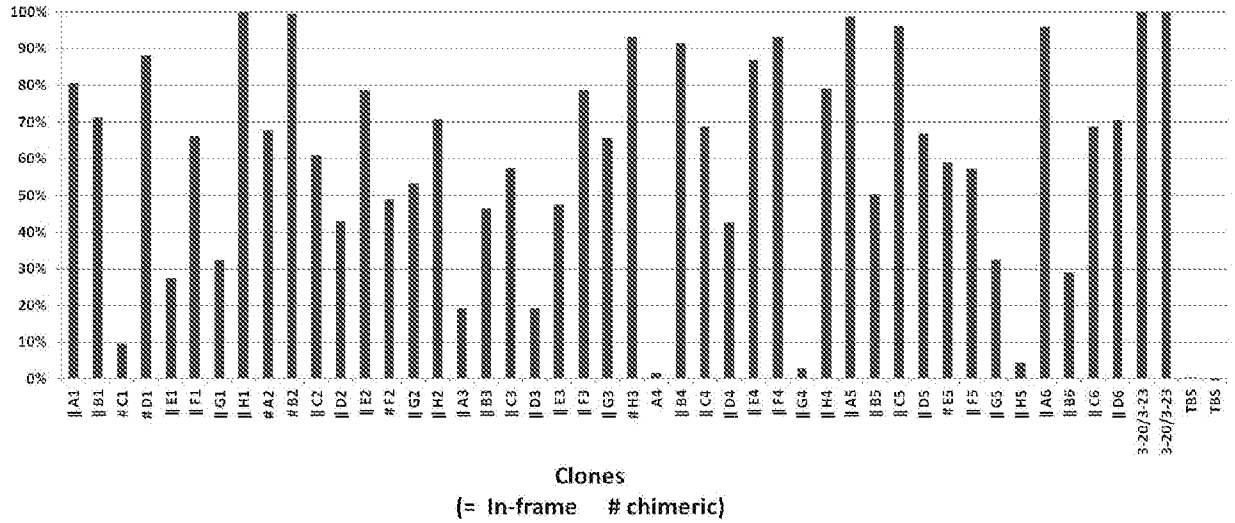
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
B1
C1
D1
E1
F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
H2
A3
B3
C3
D3
E3
F3
G3
H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
3-20/3-23
3-20/3-23
TBS
TBS
Clones
(= In-frame # chimeric)

FIGURE 15

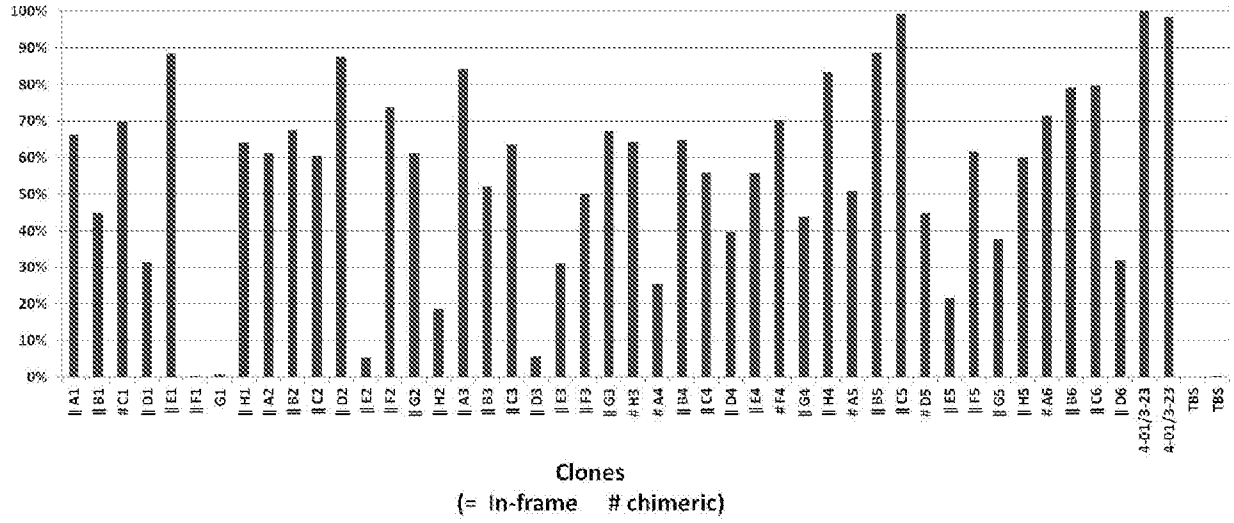
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
B1
C1
D1
E1
F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
H2
A3
B3
C3
D3
E3
F3
G3
H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
4-01/3-23
4-01/3-23
TBS
TBS
Clones
(= In-frame # chimeric)

FIGURE 16

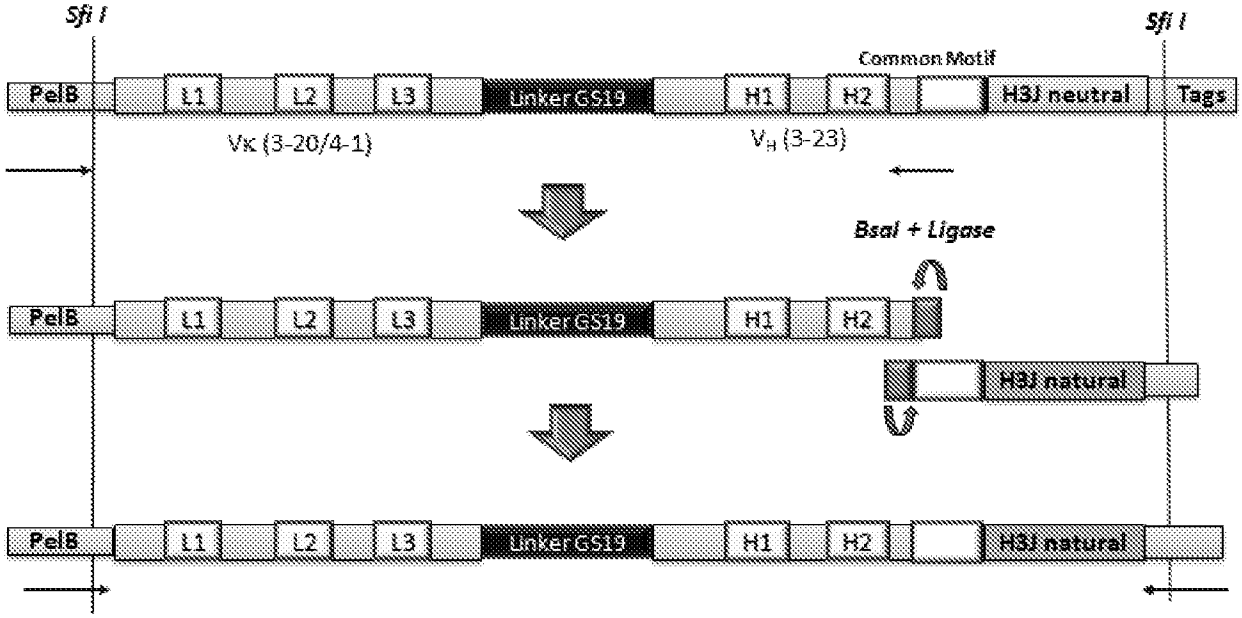
Sfi I
Sfi I
Common Motif
PelB
L1
L2
L3
Linker GS19
H1
H2
H3J neutral
Tags
Vκ (3-20/4-1)
VH (3-23)
BsaI + Ligase
PelB
L1
L2
L3
Linker GS19
H1
H2
H3J natural
PelB
L1
L2
L3
Linker GS19
H1
H2
H3J natural

FIGURE 17

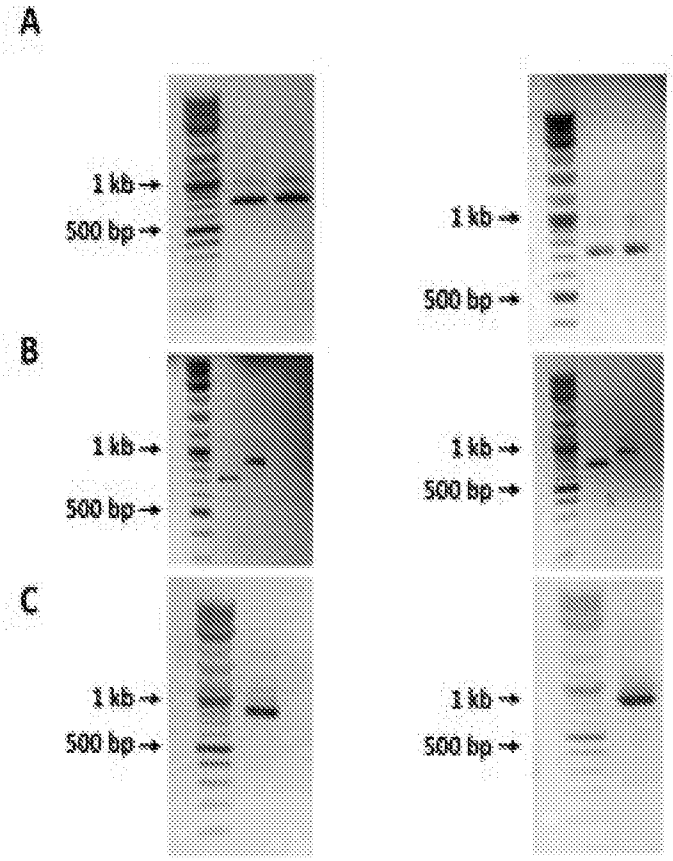
A
1 kb
500 bp
1 kb
500 bp
B
1 kb
500 bp
1 kb
500 bp
C
1 kb
500 bp
1 kb
500 bp

FIGURE 18

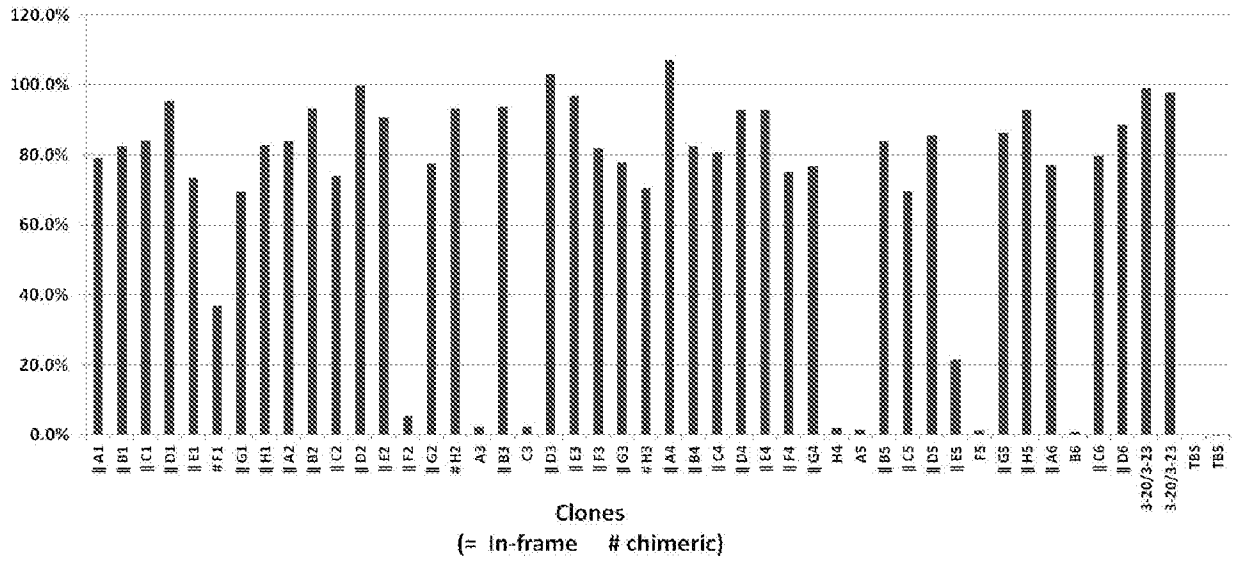
120.0%
100.0%
80.0%
60.0%
40.0%
20.0%
0.0%
A1
B1
C1
D1
E1
#F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
#H2
A3
B3
C3
D3
E3
F3
G3
#H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
3-20/3-23
3-20/3-23
TB5
TB5
Clones
(= In-frame # chimeric)

FIGURE 19

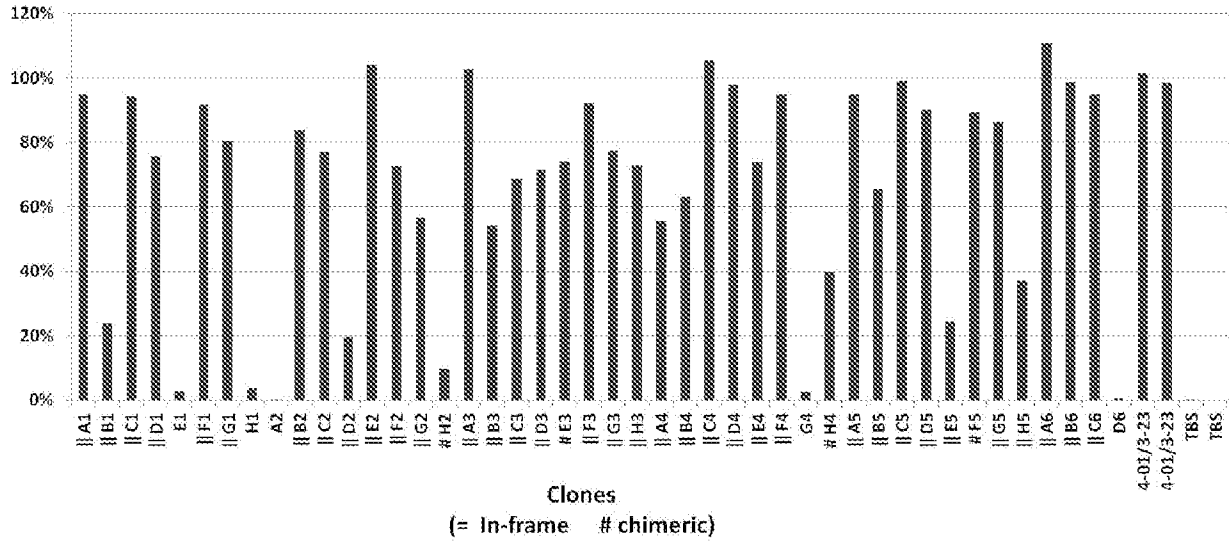
120%
100%
80%
60%
40%
20%
0%
|| A1
|| B1
|| C1
|| D1
E1
|| F1
|| G1
H1
A2
|| B2
|| C2
|| D2
|| E2
|| F2
|| G2
H2
|| A3
|| B3
|| C3
|| D3
E3
|| F3
|| G3
|| H3
|| A4
|| B4
|| C4
|| D4
|| E4
|| F4
G4
H4
|| A5
|| B5
|| C5
|| D5
|| E5
F5
|| G5
|| H5
|| A6
|| B6
|| C6
D6
4-01/3-23
4-01/3-23
TBS
TBS
Clones
(= In-frame # chimeric)

FIGURE 20

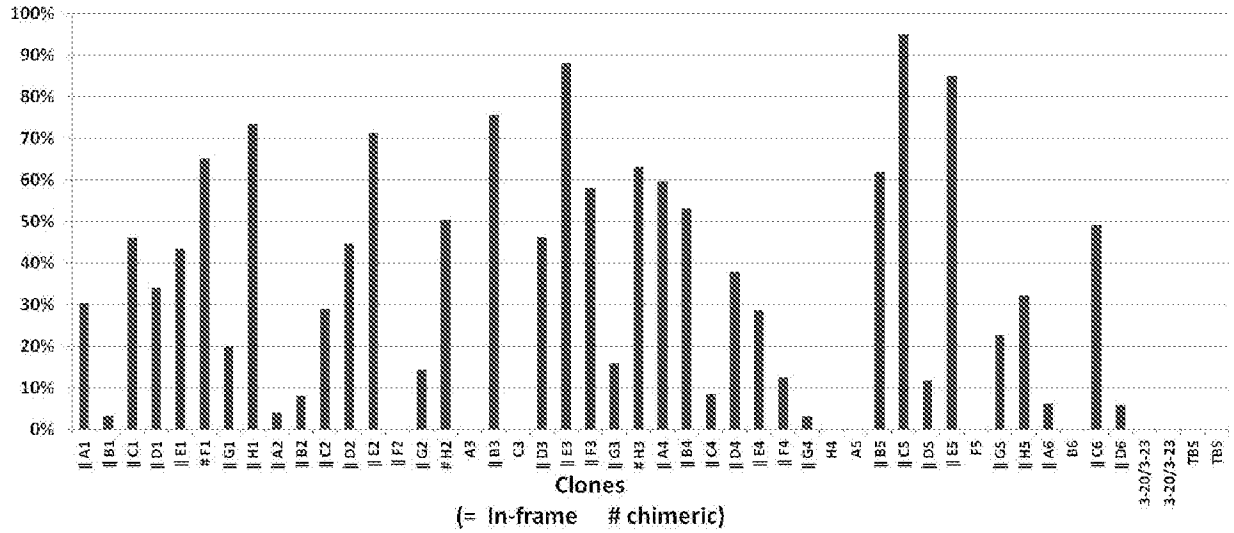
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
B1
C1
D1
E1
#F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
#H2
A3
B3
C3
D3
E3
F3
G3
#H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
3-20/3-23
3-20/3-23
TB5
TB5
Clones
(= In-frame # chimeric)

FIGURE 21

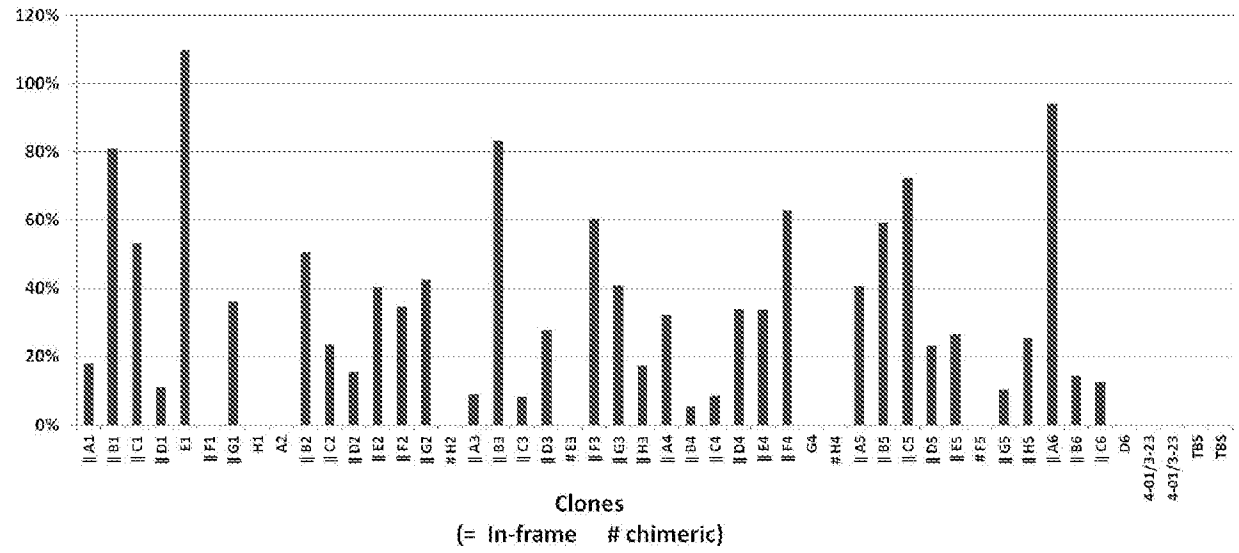
120%
100%
80%
60%
40%
20%
0%
A1
B1
C1
D1
E1
F1
G1
H1
A2
B2
C2
D2
E2
F2
G2
H2
A3
B3
C3
D3
E3
F3
G3
H3
A4
B4
C4
D4
E4
F4
G4
H4
A5
B5
C5
D5
E5
F5
G5
H5
A6
B6
C6
D6
4-01/3-23
4-01/3-23
TBS
TBS
Clones
(= In-frame # chimeric)

FIGURE 22

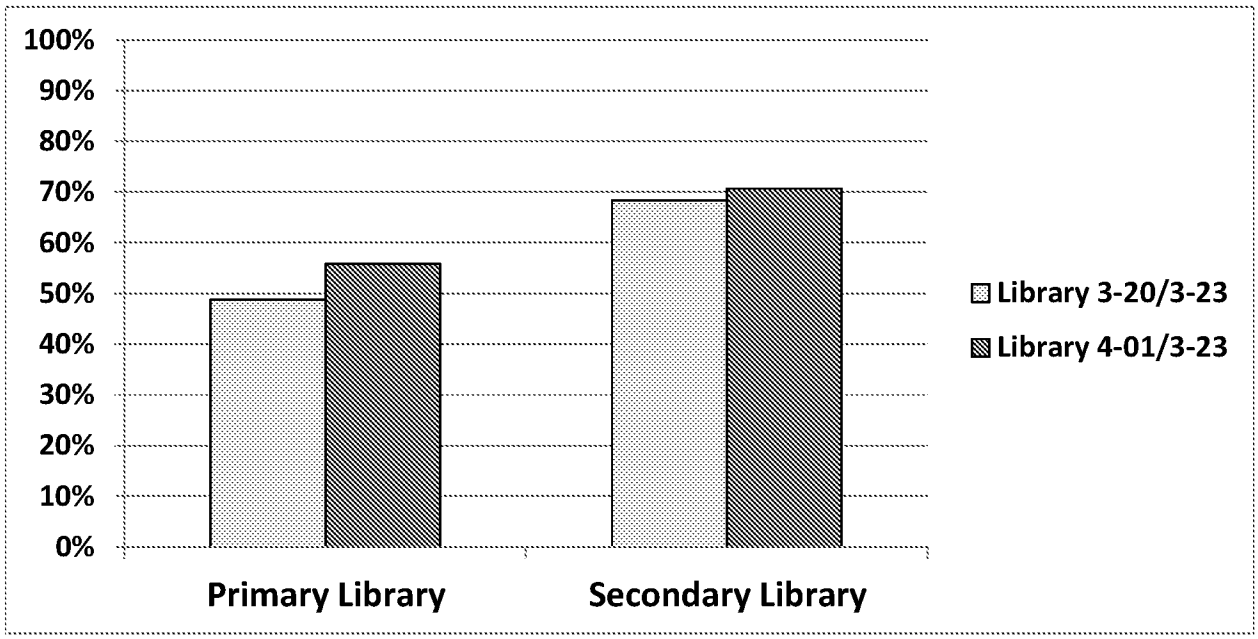
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Primary Library
Secondary Library
Library 3-20/3-23
Library 4-01/3-23

FIGURE 23