A COLLECTION AND METHODS FOR ITS USE

CROSS REFERENCE

This application claims the benefit of U.S. provisional application serial number 61/494,452 filed June 8, 2011, and U.S. provisional application serial number 61/415,367 filed November 19, 2010, which are both incorporated by reference in their entireties.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on February 15, 2011, is named MS130US.txt and is 229, 568 bytes in size.

BACKGROUND

Advances in pharmaceutical development, especially in the field of therapeutic antibodies, are rapidly enabling and/or improving the treatment of many diseases. These advances, by reaching novel target spaces and providing novel mechanisms of action are increasingly improving the quality of lives of patients even with the most severe and challenging diseases. One challenge for the health care system in general and patients in particular is that the costs of new drugs, enabled by of these pharmaceutical advances, are also rapidly increasing. The high costs are a result of the investments required for the development of pharmaceuticals, especially of antibodies, which currently exceed one billion dollars per marketed product. The high risk of failure in development and very long developmental timelines make these investments inevitable. It may take over fifteen years from the time of identification of a potential therapeutic antibody until it reaches the market and can benefit patients. Each stage of development, from identification, pre-clinical, clinical to market entry is riddled with challenges and risks. Pharmaceutical companies are constantly working to reduce developmental costs by reducing timelines and risks of failure in order to get the most effective medicines into the hands of patients quickly.

The following disclosure provides a valuable advance which allows for faster identification of the optimal therapeutic antibodies for the treatment of any disease. Therapeutic antibody candidates must fulfill a number of development criteria in order to make it to the market, such as, long term stability, low aggregation propensity and high expression yields. The disclosed advance increases the probability and speed of identifying an antibody that can fulfill all of the rigorous development criteria right from the start. The resultant antibody will be less expensive to produce and will be effective and safe in the treatment of numerous diseases.

A well known method of identifying therapeutic antibodies is through the use of phage display technology. Phage display utilizes virus-like particles that are grown in bacteria to display antibodies. One benefit of this technology is that the libraries used are massive, with up to 1 X 10^{11} antibodies, which can quickly be tested for binding to any target relevant for any disease. *See*, for example, Knappik *et al.*, (2000), "Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides," *J. Mol. Biol.* 11;296(1):57-86, and US Patent Number 6,300,064, both of which are incorporated by reference in their entireties. The benefit of working with such large numbers is that the output of a screening against a target may result in hundreds of antibodies that bind to the therapeutic target, all of which could be therapeutically relevant. A problem, though, is that often only a few of these antibodies are developable, meaning that they can meet all of the rigorous criteria required in order to make it to the market.

In order for a new phage display collection to shorten the identification timelines and reduce the inherent risks, the collection should comprise antibodies having the properties which are necessary for selection and clinical development and which will result in safe and effective treatment in patients. Such properties include: 1) high phage display rates, so that each and every antibody of the collection can be tested against the target of interest; 2) high expression levels in both Fab and IgG1 formats, so that the antibody or fragment can be reproduced efficiently with the needed quantity; 3) high thermal stability in both Fab and IgG1 formats, to ensure structural and functional integrity of the molecules delivered to patients; 4) high stability in serum in both Fab and IgG1 formats, so that the antibody shows increased half-life and prolonged activity; 5) high monomeric content (% monomer) as determined by size exclusion chromatography (SEC) in both Fab and IgG1 formats as this signifies a low aggregation propensity; 6) high isoelectric point (pI) in IgG1 format; 7) high thermal stability in Fab and IgG1 formats before and after exposure to acid; 8) low turbidity in Fab or IgG1 formats before and after exposure to acid; 9) stable molecular radius and % polydispersity before and after exposure to acid; 10) low risk of immunogenicity, thereby increasing safety, and/or 11) high diversity, so that one collection can be used to identify many antibodies against any therapeutic target.

A collection, which in essential ways imitates the human immune system, should be highly valuable, or even the optimal solution. The human immune system is composed of antibodies encoded by germline genes. Antibodies, in part, comprise of a variable heavy chain and variable light chains. There are approximately 50 variable heavy chain germline genes and approximately 50 variable light chain germline genes, combined providing about 2,500

combinations of different variable heavy and light chain pairs. In humans, all 2500 of these combinations are believed to be produced. It has been found, though, that certain variable heavy chains, variable light chains and/or variable heavy and light chain combinations (pairs) are present at a higher level than others. It was hypothesized that there must be some reason that some are present more than others, and if so, that the highly present germline genes may have favorable functional properties. Therefore, one way of providing a collection of antibodies having favorable functional properties is to generate a collection comprising the abundant variable heavy chain, variable light chain, and/or variable heavy chain and variable light chain pairs present in the human immune repertoire.

In addition, the germline gene sequences present in humans are thought to have very low immunogenicity, for obvious reasons, therefore these sequences can be imitated in recombinant antibodies in order to lower the risk of immunogenicity.

Approaches to evaluate the variable heavy and light chain germline gene pairings prevalent in the human immune repertoire have been undertaken. See de Wildt et al., Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire, J Mol Biol. 22;285(3):895-901 (January 1999), which is incorporated by reference in its entirety. Wildt et al. took blood samples from human donors, sorted the IgG+ B cells, which had undergone somatic hypermutation, PCR amplified the cDNAs, sequenced each cDNA, and aligned each sequence to the known human variable domain germline genes. Wildt et al. observed that only a few germline genes dominated the immune repertoire and that the frequent heavy and light chain gene segments are often paired.

Attempts at maintaining the heavy and light chain variable domain pairings of individual B cells have also been undertaken. For example, libraries of variable domain "cognate pairs" have been disclosed. *See* Meijer et al., Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing, J Mol Biol., 358(3):764-72 (May 5 2006); and WO2005042774, which are both incorporated by reference in their entirety. Libraries according to the techniques described in Meijer et al. have been generated from individual B cells from an immunized host. Generally, the B cells are sorted by FACS so that $CD38^{HI}$ B cells, which represent somatically hypermutated cells, are selected, their cDNAs are PCR amplified, and the antibody gene products are inserted into Fab vectors for selection. Such cognate pair libraries are not without their limitations. For example, the hosts providing the B cells typically are immunized; and the B cell populations sorted have been hypermutated, therefore, the resulting libraries are biased towards a particular immunogen.

Additionally, attempts at utilizing prominent variable heavy chain or variable light chains for collection generation have been undertaken. For example, in Shi et al., "De Novo Selection of High-Affinity Antibodies from Synthetic Fab Libraries Displayed on Phage as pIX Fusion Proteins; J Mol Biol., 397(2):385-96 (March 26, 2010) and the respective patent application WO2009085462; and WO2006014498, which are incorporated by reference in their entireties. There, variable heavy chain or variable light chain germline protein sequences were incorporated into libraries based upon their frequency of use in the human immune repertoire.

Additional attempts have also been undertaken, which incorporate a specific germline pair into a collection. For example, WO1999020749, which is incorporated by reference in its entirety, describes a collection where its members comprise heavy chains having the canonical structure of a hypervariable loop encoded by the human germline heavy chain gene segment DP-47 (IGHV3-23) and/or framework regions encoded by the germline gene, and/or light chains having the canonical structure of a hypervariable loop encoded by the human germline light chain gene segment O2/O12 (IGKV1-39/1D-39) and/or framework regions encoded by the germline gene.

Additional approaches have generated libraries directly from or derived from B cells. For example, Glanville et al., Precise Determination of the Diversity of a Combinatorial Antibody Library Gives Insight into the Human Immunoglobulin Repertoire, Proc Natl Acad Sci 1;106(48):20216-21 (December 2009), which is incorporated by reference in its entirety, which describes an antibody collection built from the diversity of 654 human donor Immunoglobulin M (IgM) repertoires. Specifically, the heavy and light chain V-gene cDNAs from 654 human donors were separately PCR amplified (separating the variable heavy and light chain pair) and the heavy and light chain domains were then randomly re-associated. WO2003052416, which is incorporated by reference in its entirety, also describes the isolation of B cells from a host exhibiting a pronounced response to a pathogen of interest, resulting from either an infection by a micro-organism or treatment with a vaccine. In WO2003052416, the cDNA encoding the CDR3 region of the variable regions was sequenced and antibody fragments comprising the dominant CDR3s were designed. WO2009100896, which is incorporated by reference in its entirety, describes the isolation of B cells from an immunized host, where the cDNAs encoding the variable heavy and light chain regions were sequenced and the abundance of the unparied variable heavy and variable light chain sequences was determined. In WO2009100896, libraries were synthesized comprising the randomly recombined variable heavy and variable light chains, wherein the antibodies were specific for one immunogen. A summary of these and

additional approaches is found in Fuh et al., Synthetic antibodies as therapeutics, Expert Opin Biol Ther.,7(1):73-87 (January 2007), which is incorporated by reference in its entirety.

There is, therefore, a high need for a collection of antibodies or fragments thereof that incorporate the variable heavy and variable light chain gene pairs present in the human immune repertoire that have favorable biophysical properties relevant to development, while at the same time excluding the pairs that exist in nature, but do not have such biophysical properties. These and other needs are satisfied by the present invention.

SUMMARY

The present disclosure provides a valuable solution to the problem of efficiently identifying antibodies or antibody fragments against any antigen that are developable and safe and effective in patients. In its most general sense, the inventors began with the idea that an antibody collection that imitates the human immune system in essential ways may be advantageous. On one level, the inventors decided to imitate the human immune system by incorporating the optimal germline gene sequences, or portions thereof, from the human immune repertoire into antibodies. As such, in some embodiments, the antibodies of the collection comprise portions, for example, framework regions that are germline in sequence. Using the germline sequences should dramatically decrease the risk of immunogenicity of recombinant antibodies for therapeutic use in patients.

In addition, the inventors worked from their hypothesis that the variable heavy chain and variable light chain germline gene pairs abundant in the human immune repertoire likely have favorable biophysical properties that would lead to more efficient clinical development and increase the safety and efficacy of the resulting antibodies in patients. As background, each B cell encodes one antibody, and each antibody comprises a variable heavy chain and variable light chain. Each of the variable heavy chain and variable light chains of an antibody can be aligned with germline sequences in order to determine the origin of the antibody, meaning from which germline gene the variable heavy chain and variable light chain are encoded. Therefore, for each antibody the variable heavy chain and variable light chain comprise a germline gene or germline protein pair, for example, VH3-23 paired with VK1-5.

In order to prove the hypothesis that the prominent germline gene pairs likely have favorable biophysical properties, the first step was to identify the variable heavy chain and variable light chain germline gene pairs prominent in the human immune repertoire. This was done by extensively searching publically available literature and by sampling B cells from a human host. As a next step, this data was pooled, analyzed and the variable heavy chain and

variable light chain germline pairs present in the human immune repertoire were ranked in terms of their prevalence. From this data it was clear that certain variable heavy chain and variable light chain germline gene pairs were present more frequently than others in the human immune repertoire.

As a next step, it had to be determined which germline protein pairs were to be tested for the functional properties relevant to development, as there are ~2500 pairs in the human immune repertoire, it is not preferred to test each one. One way would be to to test the variable heavy chain and variable light chain germline protein pairs that occur most prominently in the human immune repertoire, for example, see Table 6. One could, for example, select the top four hundred pairs for testing, or select the variable heavy chain and variable light chain germline gene pairs present at or above a certain threshold number. This approach would require the synthesis and testing of a very large number of variable heavy chain and variable light chain germline protein pair sequences; therefore, such an approach may not be very efficient.

As an alternative approach, the inventors selected a subset of the variable heavy chain and variable light chain germline pairs that are representative of, accurately reproduce, or cover the majority of the prominent pairs of the human immune repertoire. This approach was based, in part, upon the observation that a small number of variable heavy, variable κ light chain, and variable λ light chain germline genes (unpaired) are dominant in the human immune repertoire. Wildt et al. at 895-896 describes this phenomenon. Wildt et al. also states that the frequent heavy and light chain gene segments are often paired, and observed that half of the pairings sampled corresponded to only five germline pairs. Therefore, a small number of the prominent heavy and light chain germline genes (unpaired) can be combined to generate a group of heavy and light chain pairs that are representative of the human immune repertoire.

This approach was undertaken in the following way. The data showing the linked VH/VL pairs, see, e.g., Table 6, and the data identifying the presence of the unlinked VH or VL chains, see, e.g. Example 3 and Table 5, was analyzed to determine the variable heavy chain, variable κ light chain, and variable λ light chain germline genes (unpaired) that are prominent in the human immune repertoire.

As a next step the prominent variable heavy chain, variable κ light chain, and variable λ light chain germline protein sequences (unpaired) were evaluated to determine their biophysical properties relevant to development, see Example 4. The variable heavy chain, variable κ light chain, and variable λ light chain germline protein sequences were evaluated *in silico* for the following properties: (i) CDR length, (ii) isoelectric point (pI) (a preferred isoelectric point is 7.5

or above as this is should provide stability in a neutral or slightly acidic formulation buffer), (iii) potential sites for potential post translational modification sites (PTM's) (specifically, N-linked glycosylation sites (NxS or NxT) or chemical modifications such as Asp cleavage (often at a DP or DQ), (iv) Asp isomerization (DS, DG), (v) deamidation (NS, NG) which can occur *in vivo* (in serum) or upon storage in formulation buffer and lead to loss of antigen binding), (vi) the presence of Methionines in the CDRs (might be prone to oxidization when exposed to solvent), (vii) the presence of unpaired Cysteines (will form disulfide bonds with any other unpaired cysteine, thus leading to crosslinking of proteins and/or lower expression levels), (viii) deviations from germline, (ix) the presence of possible T-cell epitopes, and (x) theoretical aggregation propensity.

As a next step the variable heavy chain, variable κ light chain, and variable λ light chain germline protein sequences having favorable *in silico* biophysical characteristics were combined to form variable heavy chain and variable light chain pairs. As shown in Table 5, and Figures 2-3, generally, the top 20 VH, top 8 Vλ and top 12 Vκ were selected for synthesis, combination and subsequent functional analysis. The germline gene sequences were synthesized and then combined in order to generate 400 germline protein pairs (20VH X 20VL) that are representative of, accurately reproduce, or cover the majority of the prominent pairs from the human immune repertoire as shown in Table 6. This was done by synthesizing the variable heavy and light chain germline genes, combining them into pairs, expressing the pairs as protein (germline protein pairs) and testing each to identify their biophysical properties. The following properties were tested: (i) relative display rate on phage in the Fab format, (ii) relative expression level in the Fab format, e.g., in *E.coli*; (iii) thermal stability in the Fab format; (iv) stability in bovine or mouse serum in the Fab format; (v) relative expression level in the IgG1 format; and (vi) stability in bovine serum in the IgG1 format.

The testing of the 400 germline protein pairs for display, expression, thermal and serum stability acted as a preliminary filter to remove the germline protein pairs that, although they exist in nature, do not have biophysical properties thought to be favorable for therapeutic development. The goal was to select a sub-group of germline protein pairs having favorable developability characterictics, while at the same time maintaining a high level of diversity within a collection so that the collection can be used to identify developable candidates against any antigen. Table 12 shows ~60 bold and underlined germline protein pairs which met the thresholds of an embodiment of the disclosure. Table 12 was previously disclosed in WO2010/136598 (MorphoSys AG), which claims the benefit of 61/182,350, and 61/299,401, which are all incorporated by reference in their entireties.

Of the 400 germline protein pairs tested (results shown in Table 12), 95 were selected for further testing. Of the 95 germline protein pairs selected for further testing, some were chosen because they met the previous criteria, and it was desirable to further test them. Others were chosen, despite not meeting certain thresholds, so that these pairs could be re-evaluated. The 95 germline protein pairs shown in Figures 16-24 were synthesized, expressed, purified and then tested in both Fab and IgG1 formats for the following a) purified Fab expression in mg/L, b) purified Fab monomeric content (% monomer), c) purified Fab thermal stability, d) purified IgG1 expression in mg/L, e) purified IgG1 monomeric content (% monomer), f) purified IgG1 thermal stability, g) IgG1 isoelectric point and h) IgG1 stress testing with exposure to acid, including differential scanning fluorometry (DSF), absorption, dynamic light scattering and particle staining.

In an embodiment, the following germline protein pairs (54) were identified as having superior functional activity related to developability (data shown in Figures 16-24): VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO:

236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252). Specifically, in this embodiment, the germline protein pairs (54) had values at or above the following thresholds for each criteria: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric content of purified Fab (as described in Example 9.1.3) of at least 98%; and f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99%. Therefore, collections comprising any number of these variable heavy chain and variable light chain pairs could be used to identify developable antibodies or fragments thereof against any antigen.

As compared to Table 32 of WO2010/136598, Table 32 shows only 21 of the 54 pairs as having certain different functional properties.

Embodiments of the present disclosure include collections comprising a subset of the germline protein pairs above (36 of the 54) having superior functional activity related to developability. In one embodiment, a collection comprises synthetic antibodies or functional fragments thereof, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ

ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252). In this embodiment, the subset (36) germline protein pairs was selected from the 54 germline protein pairs based upon the stress testing data. The stress testing data was identified using the methods described in Examples 9.2.5 (a-d), data shown in Figures 19-24, Example 9.2.6 (a-d), data shown in Figures 19-54 and Example 9.2.7, scoring shown in Figures 55-60. The stress testing evaluated the 95 germline protein pairs in IgG1 format in order to determine their ability to withstand exposure to acid and agitation with glass beads. An antibody's ability to withstand exposure to acid is an increasingly important factor, as a virus inactivation step is standard during the downstream processing (DSP) of Chemistry, Manufacturing and Control (CMC). The ability of antibodies or antibody fragments to resist sheer forces is a helpful criterion as filtration steps cannot be avoided during processing and sheer forces occur during administration via syringe needles or plastic tubes.

The above subset collection, (36) germline protein pairs of an embodiment, were selected as they have additional superior functional properties relevant to developability as they showed stronger resistance to acid and agitation stress than the other pairs of the 54. The 36 germline protein pairs selected in this embodiment, had values at or above the following thresholds for each criteria: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e)

monomeric content of purified Fab (as described in Example 9.1.3) of at least 98%; f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99% and g) stress testing cumulative score (as described in Example 9.2.7) of at least 1225.

As compared to Table 32 of WO2010/136598, Table 32 shows only 14 of the 36 pairs as having certain different functional properties. Additionally, WO2010/136598 does not disclose the specific combination of the 36 pairs.

In another embodiment, the thresholds for each criterion were selected as follows: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric content of purified Fab (as described in Example 9.1.3) of at least 99%; f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99%; g) isoelectric point of purified IgG1 (as described in Example 9.2.4) of at least 8.3; and h) stress testing cumulative score (as described in Example 9.2.7) of at least 1225. In this embodiment, a collection comprises (33 pairs): VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

As compared to Table 32 of WO2010/136598, Table 32 shows only 14 of the 33 pairs as having certain different functional properties. Additionally, WO2010/136598 does not disclose the specific combination of the 33 pairs.

In a further embodiment, pairs were added to a collection even though the pairs themselves did not meet all of the thresholds within each criteria, but were added to the collections in order to enhance diversity. In an embodiment, a collection further comprises: VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); and VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256). In this embodiment, a collection comprises (36 pairs): VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

Such collections overcome many of the problems of the prior art. For example, collections derived from B cells include VH/VL pairs that do not have favorable biophysical properties, as the VH and VL pairings present in such a collection are identical to the pairings

present in the sample of B cells. If a large enough sample of B cells is taken, each of the approximately 50 VH and 50 VL class pairing combinations (2500) will be present. The extensive testing of VH and VL pairs in the present disclosure shows that many of the VH and VL germline gene pairs (germline protein pairs) that exist in nature fail to have properties that would allow for developability in the clinic. Therefore, such B cell libraries comprise many VH and VL pairs that are likely not developable. Therefore, it may be desirable to generate libraries of large diversity comprising the VH and VL pairs having advantageous functional properties, but with a B cell collection approach, this is not possible.

For example, an aspect of the present disclosure is a collection of antibodies or functional fragments comprising the variable heavy and light chain germline protein pairs having advantageous properties that enhance developability, but excluding variable heavy and light chain germline gene pairs not having such properties, even if they are prominently expressed in the human immune repertoire. In this way, the collection was designed to exclude the variable heavy and light chain combinations or pairs that occur in nature (out of the 2,500 pairs) which fail to have advantageous functional properties. For example, VH4-34 is frequently occurring in the human immune repertoire as shown in Table 5, but it is also known that antibodies derived from this heavy chain germline gene can be B cell cytotoxic, therefore, antibodies derived from this gene could be excluded from a collection design. See Bhat et al., Rapid cytotoxicity of human B lymphocytes induced by VH4-34 (VH4.21) gene-encoded monoclonal antibodies, Clin Exp Immunol.,105(1):183-90 (July 1996).

DESCRIPTION OF THE DRAWINGS

Figure 1 shows the restriction sites selected for incorporation into the C-terminus of the phoA and ompA E.coli signal sequences, as described in detail in Example 1, and includes the restriction sites around CDR 3 and their respective orientations. This figure, while displaying the E.coli signal sequences, also represents the C-terminal restriction sites selected for incorporation in the human heavy chain and kappa chain leader sequences for use in IgG1 expression, as also described in detail in Example 1.

Figure 2 shows the 20 VH germline genes selected for synthesis, combination and functional characterization, as described in detail in Example 4. The figure also shows the results of the *in silico* analysis of each germline gene, where pI represents isolelectric point, PTMs are potential

post translational modification sites in the complementarity determining regions, as described herein, NxS/T are N-linked glycosylation sites, and Met in CDR are methionines.

Figure 3 shows the 8 Vλ and 12 Vκ germline genes selected for synthesis, combination and functional characterization, as described in detail in Example 4. The figure also shows the results of the *in silico* analysis of each germline gene, where pI represents isolelectric point, PTMs are potential post translational modification sites in the complementarity determining regions, as described herein, NxS/T are N-linked glycosylation sites, and Met in CDR are methionines. Here, VL means Vλ.

Figure 4 shows the VH/Vκ pairs of the pooled data from Examples 2.1 and Example 2.2. The numerical entries represent the number of each VH/Vκ germline gene pair from an individual B cell identified in the pooled data. The Y axis shows the VH germline genes ranked from top (most prevalent) VH3-23 to bottom (less prevalent) VH3-20 in terms of frequency of expression in the pooled data. The X axis shows the Vκ germline genes ranked from left (most prevalent) IGKV3-20 to right (less prevalent) IGKV1D-17 in terms of frequency of expression in the pooled data. The number 1358 is the number of B cells sampled.

Figure 5 shows the VH/Vλ pairs of the pooled data from Examples 2.1 and Example 2.2. The numerical entries represent the number of each VH/Vλ germline gene pair from an individual B cell identified in the pooled data. The Y axis shows the VH germline genes ranked from top (most prevalent) VH3-23 to bottom (less prevalent) VH3-20 in terms of frequency of expression in the pooled data. The X axis shows the Vλ germline genes ranked from left (most prevalent) IGLV2-14 to right (less prevalent) IGLV4-60 in terms of frequency of expression in the pooled data. The number 779 is the number of B cells sampled.

Figures 6A-C show the amino acid sequences encoded by the VH germline genes (SEQ ID NOS 63-118, respectively, in order of appearance), as described in Tomlinson et al., (1992), "The Repertoire of Human Germline Vh Sequences Reveals about Fifty Groups of Vh Segments with Different Hypervariable Loop" J. Mol. Biol. 227, 776-798; Matsuda et al. (1998), "The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus" J Exp Med 188(11):2151-62; and LeFranc MP (2001) "Nomenclature of the human immunoglobulin heavy (IGH) genes." Exp Clin Immunogenet. 18(2):100-16.

Figures 7A-C show the amino acid sequences encoded by the Vκ germline genes (SEQ ID NOS 119-164, respectively, in order of appearance), as described in Schäble and Zachau

(1993), “The variable genes of the human immunoglobulin kappa locus,” Biol. Chem Hoppe Seyler. 374(11):1001-22; Brensing-Küppers et al. (1997), “The human immunoglobulin kappa locus on yeast artificial chromosomes (YACs)” Gene. 191(2):173-81; Kawasaki et al. (2001), “Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes” Eur J Immunol 31(4):1017-28; and Lefranc MP (2001) "Nomenclature of the human immunoglobulin kappa (IGK) genes" Exp Clin Immunogenet., 18, 161-174.

Figures 8A-B show the amino acid sequences encoded by the Vλ germline genes (SEQ ID NOS 165-202, respectively, in order of appearance), as described in Kawasaki et al., (1997) “One-Megabase Sequence Analysis of the Human immunoglobulin lambda Gene Locus” Genome Research 7(3):250-61; Frippiat et al., (1995) "Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2" Hum. Mol. Genet., 4, 983-991; and LeFranc MP (2001) "Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp Clin Immunogenet.;18:242–254.

Figure 9 shows the pJPd1 Fab tricistronic phage display vector.

Figure 10 shows the pJPx1 Fab expression vector.

Figure 11 shows the pMx11 (pMORPHX11) Fab expression vector.

Figure 12 shows the pMORPH30 Fab display vector.

Figure 13 shows the pJP_h_IgG1f variable heavy chain IgG1 expression vector.

Figure 14 shows the pJP_h_Ig_kappa variable κ light chain IgG expression vector.

Figure 15 shows the pJP_h_Ig_lambda2 variable λ light chain IgG expression vector.

Figure 16 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the purified Fab expression yield in mg/L (culture), purified Fab monomeric content (% monomer), purified Fab thermal stability in °C, purified IgG1 expression yield in mg/L (cell culture), purified IgG1 monomeric content (% monomer), purified IgG1 thermal stability in °C (the transition shown is that of the variable domains, the transition of the Fc domains is not shown) and IgG1 isoelectric point of the tested germline protein pairs numbers 1-32. The data was determined using the methods described in Example 9.1.1-9.1.3 and 9.2.1-9.2.4. Here, VL means Vλ.

Figure 17 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the purified Fab expression yield in mg/L (culture), purified Fab monomeric content (% monomer), purified Fab thermal stability in °C, purified IgG1 expression yield in mg/L (cell culture), purified IgG1 monomeric content (% monomer), purified IgG1 thermal stability in °C (the transition shown is that of the variable domains, the transition of the Fc domains is not shown) and IgG1 isoelectric point of the tested germline protein pairs numbers 33-64. The data was determined using the methods described in Example 9.1.1-9.1.3 and 9.2.1-9.2.4. Here, VL means Vλ.

Figure 18 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the purified Fab expression yield in mg (purified Fab)/L (culture), purified Fab monomeric content (% monomer), purified Fab thermal stability in °C, purified IgG1 expression yield in mg (purified IgG1)/L (cell culture), purified IgG1 monomeric content (% monomer), purified IgG1 thermal stability in °C (the transition shown is that of the variable domains, the transition of the Fc domains is not shown) and IgG1 isoelectric point of the tested germline protein pairs numbers 65-95. The data was determined using the methods described in Example 9.1.1-9.1.3 and 9.2.1-9.2.4. Here, VL means Vλ.

Figure 19 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the thermal stability in °C (apparent Tm) before and after acid exposure (the apparent Tm given corresponds to the unfolding of the variable domains, the unfolding midpoint of the Fc domains is not shown) as determined using differential scanning fluorometry as described in Example 9.2.5(a), the relative change in turbidity based upon the UV absorption before and during acid exposure and after neutralization as described in Example 9.2.5(b). The data shown is of the tested germline protein pairs numbers 1-32. Here, VL means Vλ.

Figure 20 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the particle radius (nm) before and after acid exposure and the polydispersity before and after acid exposure as described in Example 9.2.5(c), the particle staining before and after acid as described in Example 9.2.5(d), and the cumulative score as described in Example 9.2.7. The data shown is of the tested germline protein pairs numbers 1-32. Here, VL means Vλ.

Figure 21 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the thermal stability in °C (apparent Tm) before and after acid exposure (the apparent Tm given corresponds to the unfolding of the variable domains, the unfolding midpoint of the Fc

domains is not shown) as determined using differential scanning fluorometry as described in Example 9.2.5(a), the relative change in turbidity based upon the UV absorption before and during acid exposure and after neutralization as described in Example 9.2.5(b). The data shown is of the tested germline protein pairs numbers 33-64. Here, VL means Vλ.

Figure 22 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the particle radius (nm) before and after acid exposure and the polydispersity before and after acid exposure as described in Example 9.2.5(c), the particle staining before and after acid as described in Example 9.2.5(d), and the cumulative score as described in Example 9.2.7. The data shown is of the tested germline protein pairs numbers 33-64. Here, VL means Vλ.

Figure 23 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the thermal stability in °C (apparent Tm) before and after acid exposure (the apparent Tm given corresponds to the unfolding of the variable domains, the unfolding midpoint of the Fc domains is not shown) as determined using differential scanning fluorometry as described in Example 9.2.5(a), the relative change in turbidity based upon the UV absorption before and during acid exposure and after neutralization as described in Example 9.2.5(b). The data shown is of the tested germline protein pairs numbers 65-95. Here, VL means Vλ.

Figure 24 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the particle radius (nm) before and after acid exposure and the polydispersity before and after acid exposure as described in Example 9.2.5(c), the particle staining before and after acid as described in Example 9.2.5(d), and the cumulative score as described in Example 9.2.7. The data shown is of the tested germline protein pairs numbers 65-95. Here, VL means Vλ.

Figure 25 shows the VH germline protein (SEQ ID NOS 204-216, respectively, in order of appearance) and DNA sequences (SEQ ID NOS 217-229, respectively, in order of appearance) of the Framework 1 and HCDR1 regions of certain variable heavy chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 6, but only include the VH germline genes selected for embodiments of the collection.

Figure 26 shows the VH germline protein (SEQ ID NOS 204-216 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 217-229 (continued), respectively, in order of appearance) of the Framework 2 and HCDR2 regions of certain variable heavy chains.

The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 6, but only include the VH germline genes selected for embodiments of the collection.

Figure 27 shows the VH germline protein (SEQ ID NOS 204-216 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 217-229 (continued), respectively, in order of appearance) of the Framework 3 region of certain variable heavy chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 6, but only include the VH germline genes selected for embodiments of the collection.

Figure 28 shows the Vκ germline protein (SEQ ID NOS 230-239, respectively, in order of appearance) and DNA sequences (SEQ ID NOS 240-249, respectively, in order of appearance) of the Framework 1 and LCDR1 regions of certain variable light chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 7, but only include the Vκ germline genes selected for embodiments of the collection.

Figure 29 shows the Vκ germline protein (SEQ ID NOS 230-239 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 240-249 (continued), respectively, in order of appearance) of the Framework 2 and LCDR2 regions of certain variable light chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 7, but only include the Vκ germline genes selected for embodiments of the collection.

Figure 30 shows the Vκ germline protein (SEQ ID NOS 230-239 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 240-249 (continued), respectively, in order of appearance) of the Framework 3 region of certain variable light chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 7, but only include the Vκ germline genes selected for embodiments of the collection.

Figure 31 shows the Vλ germline protein (SEQ ID NOS 250-257, respectively, in order of appearance) and DNA sequences (SEQ ID NOS 258-265, respectively, in order of appearance) of the Framework 1 and LCDR1 regions of certain variable light chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 8, but only include the Vλ germline genes selected for embodiments of the collection. Here, VL means Vλ.

Figure 32 shows the Vλ germline protein (SEQ ID NOS 250-257 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 258-265 (continued), respectively, in order of appearance) of the Framework 2 and LCDR2 regions of certain variable light chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 8, but only include the Vλ germline genes selected for embodiments of the collection. Here, VL means Vλ.

Figure 33 shows the Vλ germline protein (SEQ ID NOS 250-257 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 258-265 (continued), respectively, in order of appearance) of the Framework 3 region of certain variable light chains. The amino acid sequences are germline protein sequences, as defined herein. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The germline genes shown are the same as those shown in Figure 8, but only include the Vλ germline genes selected for embodiments of the collection. Here, VL means Vλ.

Figure 34 shows the VH germline protein (SEQ ID NOS 266-278, respectively, in order of appearance) and DNA sequences (SEQ ID NOS 279-291, respectively, in order of appearance) of the Framework 1 and HCDR1 regions of certain variable heavy chains. The amino acid sequences have been modified within HCDR1 to remove potential post translational modification sites (PTMs). The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The amino acids that have been modified in HCDR1 are underlined and the corresponding DNA encoding each position is bold and underlined.

Figure 35 shows the VH germline protein (SEQ ID NOS 266-278 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 279-291 (continued), respectively, in order of appearance) of the Framework 2 and HCDR2 regions of certain variable heavy chains.

The amino acid sequences have been modified within HCDR2 to remove potential post translational modification sites (PTMs). The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. The amino acids that have been modified in HCDR2 are underlined and the corresponding DNA encoding each position is bold and underlined.

Figure 36 shows the VH germline protein (SEQ ID NOS 266-278 (continued), respectively, in order of appearance) and DNA sequences (SEQ ID NOS 279-291 (continued), respectively, in order of appearance) of the Framework 3 region of certain variable heavy chains. The amino acid sequences are germline as no potential post translation modification sites were removed within the framework regions. The DNA sequences have been codon optimized by GeneArt for E.coli expression avoiding rare human codons. VH1-69*01 and VH3-23 may also have nucleotides CGT at position 94.

Figure 37 shows representative antibodies or antibody fragments specific for Dkk3 identified from the sub-collections VH3-23/VK1-39, and VH3-23/VL3-1,as described in Example 11. The figure shows the sub-collection from which each antibody or fragment was identified, the antigen, the length of the CDR-H3 and CDR-L3, the Fab thermal stability and affinity, the IgG1 pI, expression yield (mg/L), thermal stability and monomeric content (% monomer) determined by SEC. Here, VL means Vλ.

Figure 38 shows representative antibodies or antibody fragments specific for ErbB4/Her4_Fc identified from the sub-collections VH3-23/VK1-39, and VH3-23/VL3-1, as described in Example 11. The figure shows the sub-collection from which each antibody or fragment was identified, the antigen, the length of the CDR-H3 and CDR-L3, the Fab thermal stability and affinity, the IgG1 pI, expression yield (mg/L), thermal stability and monomeric content (% monomer) determined by SEC. Here, VL means Vλ.

Figure 39 shows apparent temperature melting points of selected Fabs as determined by Differential Scannning Fluorimetry (DSF) as described in Example 9.1.2. Each dot represents one unique Fab. Squares indicate the control Fabs as described in Example 9. Bars indicate the Median. The control represents the antibody tested for functional properties in Example 9, comprising germline FR regions and CDR1 and 2 of the respective germline protein pair, and the CDR3 from Ewert et al. The selected Fabs were generated in Example 11, and differ in sequence from the control antibody only in the CDR3. The close clustering here, shows that the output of the collection, meaning antibodies or fragments selected against DKK3 or

ErbB4/Her4_Fc antigen, maintain the superior functional properties of the members of the collection design.

Figure 40 shows the amino acid sequences (SEQ ID NOS 293, 295, 297 and 301, respectively, in order of appearance) and codon optimized nucleic acid sequences (SEQ ID NOS 292, 294, 296, 298, 299, 300, 302 and 303, respectively, in order of appearance) encoding the FR4 regions of collections of the invention.

Figures 41A and B show the amino acid sequence (SEQ ID NO: 305) and codon optimized nucleic acid sequence (SEQ ID NO: 304) encoding the IgG1f heavy chain constant domain of collections of the invention. The nucleic acid sequences shown have been codon optimized.

Figures 42 shows the amino acid sequence (SEQ ID NO: 307) and codon optimized nucleic acid sequences (SEQ ID NO: 306) encoding the Fab heavy chain constant domain of collections of the invention.

Figure 43 shows the amino acid sequences (SEQ ID NOS 309 and 311, respectively, in order of appearance) and codon optimized nucleic acid sequences (SEQ ID NOS 308 and 310, respectively, in order of appearance) encoding the IgG1f and Fab kappa light chain constant domains of collections of the invention. The nucleic acid sequences shown have been codon optimized.

Figure 44 shows the amino acid sequences (SEQ ID NOS 313 and 315, respectively, in order of appearance) and codon optimized nucleic acid sequences (SEQ ID NOS 312 and 314, respectively, in order of appearance) encoding the IgG1f and Fab lambda light chain constant domains of collections of the invention.

Figure 45 shows isoelectric point (pI) values of selected IgGs as described in Example 9.2.4. Each dot represents one unique IgG. Squares indicate the control IgGs as described in Example 9. Bars indicate the Median. The control represents the antibody tested for functional properties in Example 9, comprising germline FR regions and CDR1 and 2 of the respective germline protein pair, and the CDR3 from Ewert et al. The selected IgGs were generated in Example 11, and differ in sequence from the control antibody only in the CDR3. The close clustering here, shows that the output of the collection, meaning antibodies or fragments selected against DKK3 or ErbB4/Her4_Fc antigen, maintain the superior functional properties of the members of the collection design.

Figure 46 shows apparent midpoints of unfolding of selected IgGs as determined by Differential Scanning Fluorimetry (DSF) as described in Example 9.2.2. Each dot represents one unique IgG. Squares indicate the control IgGs as described in Example 9. Bars indicate the Median. The control represents the antibody tested for functional properties in Example 9, comprising germline FR regions and CDR1 and 2 of the respective germline protein pair, and the CDR3 from Ewert et al. The selected IgGs were generated in Example 11, and differ in sequence from the control antibody only in the CDR3. The close clustering here, shows that the output of the collection, meaning antibodies or fragments selected against DKK3 or ErbB4/Her4_Fc antigen, maintain the superior functional properties of the members of the collection design.

Figure 47 shows expression yields of selected IgGs as determined by UV-spectrophotometry as described in Example 9.2.1. Each dot represents one unique IgG. Squares indicate the control IgGs as described in Example 9. Bars indicate the Median. The control represents the antibody tested for functional properties in Example 9, comprising germline FR regions and CDR1 and 2 of the respective germline protein pair, and the CDR3 from Ewert et al. The selected IgGs were generated in Example 11, and differ in sequence from the control antibody only in the CDR3. The close clustering here, shows that the output of the collection, meaning antibodies or fragments selected against DKK3 or ErbB4/Her4_Fc antigen, maintain the superior functional properties of the members of the collection design.

Figure 48 shows monomeric content of selected IgGs as determined by size exclusion chromatography (SEC) as described in Example 9.2.3. Each dot represents one unique IgG. Squares indicate the control IgGs as described in Example 9. Bars indicate the Median. The control represents the antibody tested for functional properties in Example 9, comprising germline FR regions and CDR1 and 2 of the respective germline protein pair, and the CDR3 from Ewert et al. The selected IgGs were generated in Example 11, and differ in sequence from the control antibody only in the CDR3. The close clustering here, shows that the output of the collection, meaning antibodies or fragments selected against DKK3 or ErbB4/Her4_Fc antigen, maintain the superior functional properties of the members of the collection design. Here, VL means Vλ.

Figure 49 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the the relative change in turbidity based upon the UV absorption before and during agitation with glass beads as described in Example 9.2.6(a). The data shown is of the tested germline protein pairs numbers 1-32. Here, VL means Vλ.

Figure 50 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the thermal stability in °C (apparent Tm) after agitation with glass beads (the apparent Tm given corresponds to the unfolding of the variable domains, the unfolding midpoint of the Fc domains is not shown) as determined using differential scanning fluorometry as described in Example 9.2.6(b) shows the particle radius (nm) after agitation with glass beads, the polydispersity after agitation with glass beads as described in Example 9.2.6(c), and the particle staining before and after agitation with glass beads as described in Example 9.2.6(d). The data shown is of the tested germline protein pairs numbers 1-32. Here, VL means Vλ.

Figure 51 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the the relative change in turbidity based upon the UV absorption before and during stress testing as described in Example 9.2.6(a). The data shown is of the tested germline protein pairs numbers 33-64. Here, VL means Vλ.

Figure 52 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the thermal stability in °C (apparent Tm) after agitation with glass beads (the apparent Tm given corresponds to the unfolding of the variable domains, the unfolding midpoint of the Fc domains is not shown) as determined using differential scanning fluorometry as described in Example 9.2.6(b) shows the particle radius (nm) after agitation with glass beads, the polydispersity after agitation with glass beads as described in Example 9.2.6(c), and the particle staining before and after agitation with glass beads as described in Example 9.2.6(d). The data shown is of the tested germline protein pairs numbers 33-64. Here, VL means Vλ.

Figure 53 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the relative change in turbidity based upon the UV absorption before and during stress testing as described in Example 9.2.6(a). The data shown is of the tested germline protein pairs numbers 65-95. Here, VL means Vλ.

Figure 54 of the 95 germline protein pairs further tested as described in Example 9, this figure shows the thermal stability in °C (apparent Tm) after agitation with glass beads (the apparent Tm given corresponds to the unfolding of the variable domains, the unfolding midpoint of the Fc domains is not shown) as determined using differential scanning fluorometry as described in Example 9.2.6(b) shows the particle radius (nm) after agitation with glass beads, the polydispersity after agitation with glass beads as described in Example 9.2.6(c), and the particle

staining before and after agitation with glass beads as described in Example 9.2.6(d). The data shown is of the tested germline protein pairs numbers 65-95. Here, VL means Vλ.

Figure 55 as described in Example 9.2.7 for each of the stress testing experiments done in Examples 9.2.5-9.2.6, exact values were identified, and for each exact value a corresponding score was provided. This figure shows the score, whether 0, 25, 75, or 100 given to each value for the experiments completed in Example 9.2.5, acid testing. The scores shown are of the tested germline protein pairs numbers 1-32. Here, VL means Vλ.

Figure 56 as described in Example 9.2.7 for each of the stress testing experiments done in Examples 9.2.5-9.2.6, exact values were identified, and for each exact value a corresponding score was provided. This figure shows the score, whether 0, 25, 75, or 100 given to each value for the experiments completed in Example 9.2.6, agitation with glass beads. In addition, this figure shows the cumulative score, which was calculated by adding together the scores from the tests done in Examples 9.2.5-9.2.6. The scores shown are of the tested germline protein pairs numbers 1-32. Here, VL means Vλ.

Figure 57 as described in Example 9.2.7 for each of the stress testing experiments done in Examples 9.2.5-9.2.6, exact values were identified, and for each exact value a corresponding score was provided. This figure shows the score, whether 0, 25, 75, or 100 given to each value for the experiments completed in Example 9.2.5, acid testing. The scores shown are of the tested germline protein pairs numbers 33-64. Here, VL means Vλ.

Figure 58 as described in Example 9.2.7 for each of the stress testing experiments done in Examples 9.2.5-9.2.6, exact values were identified, and for each exact value a corresponding score was provided. This figure shows the score, whether 0, 25, 75, or 100 given to each value for the experiments completed in Example 9.2.6, agitation with glass beads. In addition, this figure shows the cumulative score, which was calculated by adding together the scores from the tests done in Examples 9.2.5-9.2.6. The scores shown are of the tested germline protein pairs numbers 33-64. Here, VL means Vλ.

Figure 59 as described in Example 9.2.7 for each of the stress testing experiments done in Examples 9.2.5-9.2.6, exact values were identified, and for each exact value a corresponding score was provided. This figure shows the score, whether 0, 25, 75, or 100 given to each value for the experiments completed in Example 9.2.5, acid testing. The scores shown are of the tested germline protein pairs numbers 65-95. Here, VL means Vλ.

Figure 60 as described in Example 9.2.7 for each of the stress testing experiments done in Examples 9.2.5-9.2.6, exact values were identified, and for each exact value a corresponding score was provided. This figure shows the score, whether 0, 25, 75, or 100 given to each value for the experiments completed in Example 9.2.6, agitation with glass beads. In addition, this figure shows the cumulative score, which was calculated by adding together the scores from the tests done in Examples 9.2.5-9.2.6. The scores shown are of the tested germline protein pairs numbers 65-95. Here, VL means Vλ.

Figures 61 A-D germline protein pairs of embodiments of the invention were displayed on phage and selected against Frizzled-4 Fc, GFP or erbB4/Her4_Fc fusion. This figure shows the sub-collections used, the antigen selected against, the number of clones screened, ELISA positive hits, and number of unique antibodies. Here, VL means Vλ.

Figures 62A-C shows IgGs from sub-collections selected against rhErbB4/Her4_Fc fusion, rhFZD-4 Fc fusion and eGFP, as described in Example 11. The figures show the sub-collection from which each antibody was identified, the antigen, the length of the CDR-H3 and CDR-L3, the IgG1 pI, IgG1 expression yield (mg/L), IgG1 thermal stability and monomeric content (% monomer) determined by SEC. Here, VL means Vλ.

DETAILED DESCRIPTION

Definitions

To facilitate understanding of the invention, the following definitions and illustrations are provided.

"Database or readable medium" as used herein, refers to any format for storing sequence data and thus any collection of information, such as a database file, a lookup table, an Excel spreadsheet or the like. In certain embodiments the database is stored in electronic form, such as a computer readable memory device. This includes media such as a server, a client, a hard disk, a CD, a DVD, a personal digital assistant such as a Palm Pilot, a tape, a zip disk, the computer's internal ROM (read-only-memory) or the internet or worldwide web. Other media for the storage of files accessible by a computer will be obvious to one skilled in the art.

"In silico" refers to manipulations, analysis, or designs performed on a computer, but may also be likewise performed on paper or mentally.

The term “antibody” as used herein includes whole antibodies. An antibody may be polyclonal, affinity-purified polyclonal, monoclonal, human, murine or rodent, chimeric, camelid or humanized antibodies. An antibody may belong to any of the antibody classes, such as IgG, IgG1, IgG2, IgG3, IgG4, IgA (including human subclasses IgA1 and IgA2), IgD, IgE, or IgM. An “antibody” is a protein comprising at least two heavy (H) chains and two light (L) chains inter-connected by disulfide bonds.

The term antibody “fragment” or “functional fragment” as used herein includes any antigen binding fragment, such as Fab, F(ab')2, Fab’, Fv, scFv, single chains which include an Fc portion, nanobodies and other antibody like structures having scaffolds other than variable framework regions. The term “functional fragment” includes, but is not limited to any portion of an antibody, that retains the ability to bind to an antigen of interest.

As used herein, the term “affinity” refers to the strength of interaction between antibody and antigen at antigenic sites. Within each antigenic site, the variable region of the antibody interacts through non-covalent forces with an antigen at numerous sites; the more interactions, the stronger the affinity. As used herein, the term “high affinity” for an antibody or functional fragment thereof, such as an IgG antibody, refers to an antibody having a KD of 10^{-8} M or less, 10^{-9} M or less, or 10^{-10} M or less, or 10^{-11} M or less, or 10^{-12} M or less for a target antigen. However, “high affinity” binding can vary for other antibody isotypes. For example, “high affinity” binding for an IgM isotype refers to an antibody having a KD of 10^{-7} M or less, or 10^{-8} M or less.

The term “Kassoc” or “Ka”, as used herein, is intended to refer to the association rate constant of a particular antibody-antigen interaction, whereas the term “Kdis” or “Kd,” as used herein, is intended to refer to the dissociation rate constant of a particular antibody-antigen interaction. The term “KD”, as used herein, is intended to refer to the equilibrium dissociation constant, which is obtained from the ratio of Kd to Ka (*i.e.* Kd/Ka) and is expressed as a molar concentration (M). KD values for antibodies can be determined using methods well established in the art. A method for determining the KD of an antibody is by using surface plasmon resonance, or using a biosensor system such as a Biacore® system.

The term “chimeric antibody” is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding site (variable region) is linked to a constant region of a different or altered class, effector function and/or species.

The term “isotype” refers to the antibody class (e.g., IgM, IgE, IgG such as IgG1 IgG2 or IgG4) that is provided by the heavy chain constant region genes. Isotype also includes modified

versions of one of these classes, where modifications have been made to alter the Fc function, for example, to enhance or reduce effector functions or binding to Fc receptors.

The term "germline" means the nucleic acid sequence encoding antibodies or functional fragments thereof that are passed down from parent to offspring.

The term “germline protein sequence” or “germline amino acid sequence” means a) the amino acid sequence of a variable region of antibody or functional fragment thereof encoded by a germline gene; b) the amino acid sequence encoded by a modified nucleic acid sequence encoding a variable region of antibody or functional fragment thereof having the same amino acid sequence as a variable region of an antibody or functional fragment thereof encoded by a germline gene, wherein the nucleic acid sequence is modified by, for example, by codon optimization, the addition of desired restriction sites, optimized GC content, the removal of undesired mRNA splice sites or the removal of mRNA instability motifs, or c) an amino acid sequence encoded by a germline gene, but with minor mutations in the amino acid sequence, such as, for the purpose of removing of an undesired cysteine, or introduction of desired restriction site, e.g. BbsI, or that result from errors in synthesis, amplification or cloning. Examples of “germline protein sequences” or “germline amino acid sequences” are shown in Figures 6-8 and 25-33. Additionally, “germline protein sequence” or “germline amino acid sequence” include the constructs as prepared in Example 5, which comprise

a) for VH: leader sequence (modified phoA incorporating a NheI RE site as shown in Table 1); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BssHII RE site (GCGCGC) as shown in Fig. 1); CDR-H3 (WGGDGFYAMDY) (SEQ ID NO: 1) of the 4D5 antibody as used in Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the JH4 FR4 (incorporating a XhoI RE site (CTCGAG) as shown in Fig. 1);

b) for Vk: leader sequence (ompA incorporating the NdeI RE site as shown in Table 2); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BbsI RE site (GAAGAC) as shown in Fig. 1), kappa-like CDR-L3 (QQHYTTPPT) (SEQ ID NO: 2) according to Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the Jk1 FR4 (incorporating a KpnI/Acc65I RE site (GGTACC) as shown in Fig. 1); and

c) for Vλ: leader sequence (ompA incorporating the NdeI RE site as shown in Table 2); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BbsI RE site (GAAGAC) as shown in Fig. 1), lambda-like CDR-L3 (QSYDSSLSGVV) (SEQ ID NO: 3) according to Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the Jl2/3 FR4 (incorporating a KpnI/Acc65I RE site (GGTACC) as shown in Fig. 1).

The “germline protein sequences” or “germline amino acid sequences” of antibodies encoded by the germline genes are disclosed in the following publications, for VH: Tomlinson et al., (1992), “The Repertoire of Human Germline Vh Sequences Reveals about Fifty Groups of Vh Segments with Different Hypervariable Loop” J. Mol. Biol. 227, 776-798; Matsuda et al. (1998), “The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus” J Exp Med 188(11):2151-62; and LeFranc MP (2001) “Nomenclature of the human immunoglobulin heavy (IGH) genes.” Exp Clin Immunogenet. 18(2):100-16; for Vλ: Kawasaki et al., (1997) “One-Megabase Sequence Analysis of the Human immunoglobulin lambda Gene Locus” Genome Research 7(3):250-61; Frippiat et al., (1995) "Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2" Hum. Mol. Genet., 4, 983-991; and LeFranc MP (2001) "Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp Clin Immunogenet.;18:242–254; and for Vκ: Schäble and Zachau (1993), “The variable genes of the human immunoglobulin kappa locus,” Biol. Chem Hoppe Seyler. 374(11):1001-22; Brensing-Küppers et al. (1997), “The human immunoglobulin kappa locus on yeast artificial chromosomes (YACs)” Gene. 191(2):173-81; Kawasaki et al. (2001), “Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes” Eur J Immunol 31(4):1017-28; and Lefranc MP (2001) "Nomenclature of the human immunoglobulin kappa (IGK) genes" Exp Clin Immunogenet., 18, 161-174, which are all hereby incorporated by reference in their entireties.

In parts of the specification, e.g. Figure 5, the nomenclature of the variable domain germline genes used within the present application are IMGT, as described in the LeFranc et al. publications cited in the previous paragraph. Regarding nomenclature, “VH” and “IGHV” mean heavy chain variable domain, wherein the numbering of the genes is IMGT; “VL”, “Vλ” and “IGLV” mean lambda light chain variable domain, wherein the numbering of the genes is IMGT and “Vκ,” “VK” and “IGKV” mean kappa light chain variable domain, wherein the numbering of the genes is IMGT. Alternatively, “VL” can be used to mean variable light chain, including Vκ and Vλ.

The term “germline gene sequence” means a) the nucleic acid sequence of a germline gene encoding a variable region of an antibody or functional fragment thereof, or b) a modified nucleic acid sequence encoding a variable region of an antibody or functional fragment thereof having the same amino acid sequence as a variable region of an antibody encoded by a germline gene, wherein the nucleic acid sequence is modified by, for example, codon optimization, the addition of desired restriction sites, optimized GC content, the removal of undesired splice sites or the removal of mRNA instability motifs.

The term “germline gene pair(s)” means the pair of nucleic acid sequences, and their corresponding germline gene, encoding a variable heavy chain and a variable light chain of an antibody or functional fragment thereof. For example, a germline gene pair could be VH3-23/Vκ1-5, where the antibody encoded by VH3-23/Vκ1-5 comprises a variable heavy chain, or a portion thereof, encoded by germline gene VH3-23 and a variable light chain, or portion thereof, encoded by germline gene Vκ1-5.

The term “germline protein pair” means an antibody or functional fragment thereof, wherein the variable heavy chain, or portion thereof, and the variable light chain, or portion thereof, a) are each encoded by a specific germline gene, or b) are each encoded by a modified nucleic acid sequence encoding a variable region of an antibody or functional fragment thereof having the same amino acid sequence as a variable region of an antibody encoded by the specific germline gene, wherein the nucleic acid sequence is modified by, for example, by codon optimization, the addition of desired restriction sites, optimized GC content, the removal of undesired mRNA splice sites or the removal of mRNA instability motifs, or c) each comprise an amino acid sequence encoded by a germline gene, but with point mutations in the amino acid sequence, such as, for the purpose of removing of an undesired cysteine, or introduction of desired restriction sites, e.g. BbsI, or that result from errors in synthesis, amplification or cloning. For example, a germline protein pair could be the antibody or functional fragment encoded by VH3-23/Vκ1-5, where the antibody comprises a variable heavy chain, or a portion thereof, encoded by germline gene VH3-23 and a variable light chain, or portion thereof, encoded by germline gene Vκ1-5. A “germline protein pair” includes the constructs as prepared in Example 5, which comprise

a) for VH: leader sequence (modified phoA incorporating a NheI RE site as shown in Table 1); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BssHII RE site (GCGCGC) as shown in Fig. 1); CDR-H3 (WGGDGFYAMDY) (SEQ ID NO: 1) of the 4D5 antibody as used in Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the JH4 FR4 (incorporating a XhoI RE site (CTCGAG) as shown in Fig. 1);

b) for Vk: leader sequence (ompA incorporating the NdeI RE site as shown in Table 2); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BbsI RE site (GAAGAC) as shown in Fig. 1), kappa-like CDR-L3 (QQHYTTPPT) (SEQ ID NO: 2) according to Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the Jk1 FR4 (incorporating a KpnI/Acc65I RE site (GGTACC) as shown in Fig. 1); and

c) for Vλ: leader sequence (ompA incorporating the NdeI RE site as shown in Table 2); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BbsI RE site (GAAGAC) as shown

in Fig. 1), lambda-like CDR-L3 (QSYDSSLSGVV) (SEQ ID NO: 3) according to Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the Jl2/3 FR4 (incorporating a KpnI/Acc65I RE site (GGTACC) as shown in Fig. 1).

The term “variable heavy chain and variable light chain pair” or “VH/VL pair” means the combination of one variable heavy chain and one variable light chain. An antibody and functional fragment, e.g. a Fab, comprises at least one variable heavy chain bound to a variable light chain, which form the antigen binding region. An example, of a variable heavy chain and variable light chain pair is the antibody or functional fragment, or portion thereof, comprising germline amino acid sequences from VH3-23/Vκ1-5, or encoded by the germline genes VH3-23/Vκ1-5, where the antibody comprises a variable heavy chain, or a portion thereof, comprising germline amino acid sequences from VH3-23, or encoded by germline gene VH3-23 and a variable light chain, or portion thereof, comprising germline amino acid sequences from Vκ1-5, or encoded by germline gene Vκ1-5.

The term “substantially all” means at least 90%. For example, substantially all of the antibodies or functional fragments comprise variable heavy chain and variable light chain framework regions comprising germline amino acid sequences of a germline protein pair having certain properties, means that at least 90% of the antibodies or fragments comprise, variable heavy chain and variable light chain framework regions comprising germline amino acid sequences of a germline protein pair having such properties.

The sequences of the JH4 for variable heavy chain, Jκ1 for variable κ light chain, and Jλ2/3 for variable λ light chain regions are described in the following publications: Scaviner et al., (1999), " Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions" Exp Clin Immunogenet. 16(4):234-40; for JH: Ravetch et al., (1981), “Structure of the human immunoglobulin mu locus: characterization of embryonic and rearranged J and D genes.” Cell 27 (3 pt 2): 583-91; for JK: Hieter et al. (1982), “Evolution of human immunoglobulin kappa J region genes.” J Biol Chem 257(3):1516-22; for JL: Kawasaki et al., (1997) “One-Megabase Sequence Analysis of the Human immunoglobulin lambda Gene Locus” Genome Research 7(3):250-61, which are all incorporated by reference herein in their entireties. The JH4 amino acid sequence is (YFDYWGQGTLVTVSS) (SEQ ID NO: 4); the Jκ1 amino acid sequence is (WTFGQGTKVEIK) (SEQ ID NO: 5); and the Jλ2/3 amino acid sequence is (VVFGGGTKLTVL) (SEQ ID NO: 6).

The term “variable domain/region/ (VH or VL)” means the region of an immunoglobulin that comprises one or more Ig domains substantially encoded by any of the VL (including Vk and Vλ), VH, JL (including Jk and Jλ), and JH nucleic acids that make up the light chain

(including κ and λ) and heavy chain immunoglobulin genetic loci respectively. A light or heavy chain variable region (VL and VH) is made up of a "framework" or "FR" region interspersed by three hypervariable regions referred to as "complementarity determining regions" or "CDRs." The extent of the framework region and CDRs have been defined using at least the following conventions: *see* Kabat, 1991, *J. Immunol.,* 147, 915-920; Chothia & Lesk, 1987, *J. Mol. Biol.* 196: 901-917; Chothia *et al*., 1989, *Nature* 342: 877-883; Al-Lazikani *et al*., 1997, *J. Mol. Biol.* 273: 927-948); see also http://www.bioc.uzh.ch/antibody/Numbering/NumFrame.html (which shows the well known numbering conventions of antibody amino acids and the location of the CDRs and framework regions), and that used in Figures 25-36.

The term "framework region" means the part of the variable domain which serves as a scaffold for the antigen binding loops. Examples of the framework regions include FR1, FR2, FR3, and FR4 of either the variable heavy or variable light chains.

The term "complementarity determining region" or "CDR" means an antibody's antigen binding loops. Each of the two variable domains of an antibody Fv fragment contains three CDRs. The complementarity determining regions include CDR1, CDR2, and CDR3 of either the variable heavy or variable light chains.

The term "human immune repertoire" means a repertoire of the nucleic acids isolated from B cells from the immune system of a human. A repertoire may be that of an individual, or a population, and may come from naïve B cells and/or antigen experienced B cells. The present invention is amenable to the determination of an immune repertoire from a single individual, provided sufficient B-cells are obtained. Preferably, the immune repertoire is obtained from multiple individuals to avoid sample biases. An example of a human immune repertoire is described in Examples 2-3.

An "antigen" and "immunogen" are defined as any molecule that is bound specifically by an antibody.

The term "specific for an antigen/immunogen" means the specific association between an antibody and a corresponding molecule. Specificity can be determined by the methods described in Example 11, such as ELISA and/or Biacore.

"CDR diversification" or "diversified CDR" is obtained by varying the amino acid composition within a CDR. A diversified CDR can be found in a collection of antibodies or fragments having one or more identical framework regions, e.g. germline framework regions, wherein the antibodies or fragments have CDR3s comprising different amino acid sequences. Diversified CDRs can be achieved by any methods known to one of skill in the art, including the methods described by the following: WO9708320, US Patent No. 6,300,064, which is

incorporated by reference in its entirety; WO2008053275, US 12/158,181, which is incorporated by reference in its entirety; WO07056441, US60/806,602, which is incorporated by reference in its entirety; WO2009036379, US 60/993,785, which is incorporated by reference in its entirety; WO2009114815, 12/922,153, which is incorporated by reference in its entirety; WO020617071, US12/762,051, which is incorporated by reference in its entirety. CDRs are generally known to be the immunogen binding regions, therefore having collections comprising members representing a large diversity within the CDRs, especially CDR3, increases the possibility that a collection will comprise antibodies or fragments thereof having specificity, and optimal properties for any immunogen.

The term “variant” means an antibody or fragment having a different amino acid sequence than another antibody or fragment. The term “variant” includes antibodies or fragments that are essentially identical in sequence in the framework regions, but have different amino acid sequences in a CDR region, e.g. CDR3. Variants of a variable heavy chain and variable light chain pair, have essentially the same amino acid sequence within the framework regions, but have different amino acid sequences within the CDR3 region.

The term “synthesis” or “synthesized” means gene synthesis, where nucleic acid sequences are synthesized into physical DNA, comprising polynucleotides. Standard DNA synthesis comprises single nucleotide synthesis, where single-stranded oligo-nucleotides are generated and then the overlapping oligonucleotides are ligated using a PCR-like assembly. Companies, such as, Sloning (Puchheim, Germany), Geneart (Regensburg, Germany), DNA2.0 (Menlo Park, CA USA), Entelechon (Regensburg, Germany), and Genscript (Piscataway, NJ USA) provide gene synthesis technology. Sloning, for example, utilizes a set of pre-made double stranded triplet nucleotides.

The term “synthetic” describes a molecule that is made outside of the human body by synthesis or synthesized, e.g. DNA. The term “synthetic” also describes a protein, e.g. antibody or fragment that is translated from a synthetic DNA molecule.

The term “collection” or “library” means at least two members. The term “member” includes, but is not limited to nucleic acids encoding antibodies or fragments thereof or the antibodies or fragments thereof themselves.

The term “nucleic acid” is used herein interchangeably with the term “polynucleotide” or “DNA” and refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference

nucleic acid. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, and peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, as detailed below, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081, 1991; Ohtsuka *et al.*, *J. Biol. Chem*. 260:2605-2608, 1985; and Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98, 1994).

As used herein, the term, “codon optimized” or “codon optimization” means that a nucleotide sequence has been altered so that it includes codons that are preferred in a certain production system, e.g. cell or organism. The optimized nucleotide sequence is engineered to retain the amino acid sequence originally encoded by the starting nucleotide sequence. In addition the nucleotide sequence may be designed to be completely or as much as possible devoid of inhibitory motifs, mRNA splice sites, mRNA instability motifs and undesired restriction sites. It can also be optimized for GC content, desired restriction sites and other parameters. Sequences may be optimized for expression in different hosts, including bacterial or eukaryotic cells, specifically mammalian cells. The amino acid sequences encoded by optimized nucleotide sequences may also be referred to as optimized.

The term “amino acid” refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an alpha carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

The terms “polypeptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues.

The terms “identical” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same.

The term “vector” refers to a polynucleotide molecule capable of transporting another polynucleotide to which it has been linked. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of nucleic acids to which they are operatively linked are referred to herein as "expression vectors." One type of vector is a “plasmid”, which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and mammalian vectors). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “recombinant expression vectors” (or simply, “expression vectors”). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. Vectors may be compatible with prokaryotic or eukaryotic cells. In the present specification, “plasmid” and “vector” may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

Vectors typically include a prokaryotic replicon which may include a prokaryotic promoter capable of directing the expression (transcription and translation) of the VH- and/or VL-coding homologs in a bacterial host cell, such as *Escherichia coli* transformed therewith. Additionally, vectors include IgG expression vectors for use in mammalian cells, e.g. see Figures 13-15. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenience restriction sites for insertion of a DNA segment. Examples of such vector plasmids include pUC8, pUC9, pBR322, and pBR329, pPL and pKK223, available commercially.

A “display vector” includes a DNA sequence having the ability to direct replication and maintenance of the recombinant DNA molecule extra chromosomally in a host cell, such as a bacterial host cell, transformed therewith. Such DNA sequences are well known in the art. Display vectors can for example be phage vectors or phagemid vectors originating from the

class of fd, M13, or fl filamentous bacteriophage. Such vectors are capable of facilitating the display of a protein including, for example, a binding protein or a fragment thereof, on the surface of a filamentous bacteriophage. Display vectors suitable for display on phage, ribosomes, DNA, bacterial cells or eukaryotic cells, for example yeast or mammalian cells are also known in the art, for example, as are viral vectors or vectors encoding chimeric proteins.

The term “recombinant host cell” (or simply “host cell”) refers to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term “host cell” as used herein. Typical host cells are prokaryotic (such as bacterial, including but not limited to *E. coli*) or eukaryotic (which includes yeast, mammalian cells, and more). Bacterial cells are preferred prokaryotic host cells and typically are a strain of Escherichia coli (E. coli) such as, for example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, Md. Preferred eukaryotic host cells include yeast and mammalian cells including murine and rodents, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line, for example HKB11 cells, PERC.6 cells, or CHO cells.

The introduction of vectors into host cells may be accomplished by a number of transformation or transfection methods known to those skilled in the art, including calcium phosphate precipitation, electroporation, microinjection, liposome fusion, RBC ghost fusion, protoplast fusion, viral infection and the like. The production of monoclonal full-length antibodies, Fab fragments, Fv fragments and scFv fragments is well known.

Transformation of appropriate cell hosts with a recombinant DNA molecule is accomplished by methods that typically depend on the type of vector and cells used. With regard to transformation of prokaryotic host cells, see, for example, Cohen et al., Proceedings National Academy of Science, USA, Vol. 69, P. 2110 (1972); and Maniatis et al., Molecular Cloning, a Laboratory Manual, Cold spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1982). With regard to the transformation of vertebrate cells with retroviral vectors containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virol., 52:456 (1973); and Wigler et al., Proceedings National Academy of Sciences, USA, Vol. 76, P. 1373-1376 (1979).

eGFP (enhanced green fluorescent protein) has the following amino acid sequence:

*MSGSHHHHHHGT*MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATY GKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQ ERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYI MADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMVLLEFVTAAGITLGMDELYKDI. The amino acids underlined and in italics represent the His tag, and those only underlined represent the addition of a restriction enzyme recognition sequence.

Collections of Antibodies or Fragments thereof

The present disclosure enables collections of antibodies or functional fragments thereof and the nucleic acids encoding such antibodies or fragments that can be used in the identification of therapeutic antibodies against any target, where the antibodies or fragments are clinically developable, safe and effective in patients. As background, the inventors assumed that the variable heavy chain and variable light chain germline gene pairs abundant in the human immune repertoire (such as, VH3-23/VK1-5) likely have favorable biophysical properties that would lead to more efficient development and increase the safety and efficacy of the resulting antibodies in patients. Such favorable biophysical properties could include: a) high relative display rate in Fab format; b) high relative Fab expression yield; c) temperature stability in both Fab and IgG format; d) bovine/mouse serum stability of both Fab and IgG format; e) high IgG1 expression yield; e) SEC monomeric content (% monomer) in both Fab and IgG format; and/or f) high IgG1 isoelectric point (pI).

Each B cell encodes one antibody, and each antibody comprises a variable heavy chain and variable light chain. Each of the variable heavy chain and variable light chains of an antibody can be aligned with a germline gene sequence (or germline protein sequence) in order to determine the origin of the antibody, meaning from which germline gene the variable heavy chain and variable light chain were derived. Therefore, for each antibody, it can be said, that the variable heavy chain and variable light chain comprise a germline gene pair, or germline protein pair, for example, VH3-23 paired with VK1-5.

In order to prove the hypothesis that the germline protein pairs abundant in the human immune repertoire likely have favorable biophysical properties, the first step was to identify the variable heavy chain and variable light chain germline gene pairs (germline protein pairs) present in the human immune repertoire. In some aspects the data is obtained from publically available literature or databases and from the sampling of B cells.

The following articles were identified and analyzed in detail: Wardemann H. et al. (2003) Science 301, 1374-1377 and any supporting tables; Yurasov S. et al. (2005) J. Exp. Med. 201, 703-712 and any supporting tables; Tsuiji M. et al. (2006) J. Exp. Med. 203, 393-401 and any supporting tables; Yurasov S. et al. (2006) J. Exp. Med. 203, 2255-2262 and any supporting tables, Tiller T. et al. (2007) Immunity 26, 205-213 and any supporting tables, and Mietzner B. et al. (2008) PNAS 105, 9727-9732 and any supporting tables, all of which are incorporated by reference in their entireties.

Alternatively, databases, such as NCBI, can be searched using Ig-Blast. As of 2005 the database contained at least 25,000 rearranged human antibody sequences in FASTA format. Of the 22,500 entries, 13,235 represented VH sequences, 1,506 represented Vκ and 2,259 represented Vλ.

Generally, in the relevant publically available literature and databases, the following methods were followed: B cells were isolated from human donors, the B cells were sorted in order to determine their stage of development or differentiation, cDNAs were generated and amplified representing the DNA encoding the antibody from each B cell, the cDNAs were sequenced, cDNAs encoding the variable heavy chain and variable light chains were aligned to the known germline gene sequences, and the germline gene pair from each B cell was determined.

In some embodiments the data was obtained from the sampling and isolation of human B cells, which comprised a method similar to that used in the literature. In these aspects the method of producing a collection of synthetic antibodies or functional fragments thereof comprises the step of obtaining data comprising the variable heavy chain and variable light chain germline gene pairs present in the human immune repertoire; wherein the obtaining step further comprises the steps of aa) isolating human B cells from a sample; ab) generating cDNA from the B cells; ac) PCR amplifying the cDNA from the B cells; ad) sequencing the PCR products; and ae) identifying the germline genes of the PCR products. Both sets of data provided the variable heavy chain and variable light chain germline gene pairs that are present in the human immune repertoire.

Using antibody sequence data, one of skill in the art, can identify the germline families and/or genes of each VH, Vκ and Vλ variable domain. Using this approach, the prominence of each VH and VL germline family and/or gene, and/or the germline family and/or gene of each VH and VL domain pair can readily be determined by one of skill in the art.

The raw data obtained from literature and from B cells was pooled, analyzed and the variable heavy chain and variable light chain germline gene pairs present in the human immune

repertoire were ranked in terms of number of each. From this data it was clear that certain variable heavy chain and variable light chain germline gene pairs are present more frequently than others in the human immune repertoire. These prominent pairs were expected to have superior biophysical properties.

As a next step, it had to be determined which germline protein pairs were to be tested for functional properties relevant to developability, as there are ~2500 pairs in the human immune repertoire. One way would be to test the variable heavy chain and variable light chain germline protein pairs that occur most prominently in the human immune repertoire, for example see Table 6. One could, for example, select the top four hundred pairs for testing, or select the variable heavy chain and variable light chain germline protein pairs present above a certain threshold number. This approach, however, would require the synthesis and testing of a large number of variable heavy chain and variable light chain germline protein pair sequences; therefore, such an approach would not be very efficient.

As an alternative approach, the inventors selected a subset of the variable heavy chain and variable light chain germline pairs that are representative of, accurately reproduce, or cover the majority of the prominent pairs from the human immune repertoire. This approach was based, in part, upon the observation that a small number of variable heavy, variable κ light chain, and variable λ light chain germline genes are dominant in the human immune repertoire. Wildt et al. at 895-896 describes this phenomenon. Wildt et al. also states that the frequently expressed heavy and light chain gene segments are often paired, and observed that half of the pairings sampled correspond to only five germline pairs. Therefore, a small number of the prominent heavy and light chain germline genes (unpaired) can be combined to generate a group of pairs that are representative of the human immune repertoire.

Therefore, the raw data was analyzed to determine the variable heavy chain, variable κ light chain, and variable λ light chain (unpaired) germline genes prominent in the human immune repertoire. The prominent variable heavy chain, variable κ light chain, and variable λ light chain germline protein sequences were then evaluated to determine their biophysical properties relevant to development. The variable heavy chain, variable κ light chain, and variable λ light chain germline protein sequences were evaluated *in silico* for the following properties: CDR length, isoelectric point (pI) the preferred isoelectric point is 7.5 or above as this is should provide stability in a standard pH 5.5 to pH 7 formulation buffer, sites of potential post translational modification sites in the complementarity determining regions (PTM's) (specifically, N-linked glycosylation sites (NxS or NxT) or chemical modifications such as Asp cleavage (often at a DP), Asp isomerization (DS, DG),, deamidation (NS, NG) which can occur

in vivo (in serum) or upon storage in formulation buffer and lead to loss of antibody binding), the presence of Methionines in the CDRs (can be oxidized when exposed to solvent), the presence of unpaired Cysteines (will form disulfide bonds with any other unpaired cysteine, thus leading to crosslinking of proteins and/or lower expression levels), deviations from germline, the presence of potential T-cell epitopes, and theoretical aggregation propensity.

As shown in Tables 5, and Figures 2 and 3, generally, the top 20 VH, top 8 Vλ and top 12 Vκ were selected for synthesis, combination and subsequent functional analysis. The germline gene sequences were synthesized and then combined in order to generate 400 germline protein pairs that are representative of the germline gene pairs found in the immune repertoire, wherein each of the variable regions has favorable biophysical properties as identified *in silico*. The 400 VH/VL germline protein pairs were tested for the following properties: a) relative display after phage production and phage ELISA in Fab format; b) relative Fab expression yield after Fab production in E. coli, E. coli cell lysis and ELISA detection of produced Fab; c) temperature stability of Fab after Fab production in E. coli, E. coli cell lysis and ELISA detection of non-denatured Fab after incubation at increased temperatures; d) bovine/mouse serum stability of Fab from E. coli lysates by ELISA detection of non-denatured Fab after incubation in bovine/mouse serum; e) relative human IgG1 expression yield levels after IgG1 production in mammalian cells and ELISA detection of secreted IgG1 from cell culture supernatants; and f) bovine serum stability of human IgG1 by ELISA detection of non-denatured Fab after incubation in bovine/mouse serum.

Of the 400 germline protein pairs tested (results shown in Table 12), 95 were selected for further testing. After synthesis, expression and purification, the 95 germline protein pairs shown in Figures 16-24 were tested in both Fab and IgG1 formats for the following a) purified Fab expression yield in mg/L, b) purified Fab monomeric content (% monomer), c) purified Fab thermal stability, d) purified IgG1 expression yield in mg/L, e) purified IgG1 monomeric content (% monomer), f) purified IgG1 thermal stability, g) IgG1 isoelectric point and h) IgG1 stress testing with exposure to acid, including differential scanning fluorometry (DSF), absorption, dynamic light scattering and particle staining. The results are shown in Figures 16-24.

In an embodiment, the following thresholds were set i) an expression yield in Fab format of at least 2.5 mg/L; ii) thermal stability at 70°C or above in Fab format; iii) monomeric content (% monomer) in Fab format of at least 98% as determined by SEC; iv) an expression yield in IgG1 format of at least 30 mg/L; v) thermal stability at 73°C or above in IgG1 format; and vii) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC. Therefore, in an embodiment, a collection comprises synthetic antibodies or functional

fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain framework regions comprising germline protein sequences of a germline protein pair, wherein said germline protein pairs comprise the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 98% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format; and

vi) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC.

In additional embodiments, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90%, or at least 95%, or each of the antibodies or functional fragments comprise variable heavy chain and variable light chain framework regions comprising germline protein sequences of a germline protein pair, wherein said germline protein pairs comprise the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 98% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format; and

vi) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC.

In additional embodiments, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein wherein the antibodies or functional fragments consists of or consists essentially of variable heavy chain and variable light chain framework regions comprising germline protein sequences of a germline protein pair, wherein said germline protein pairs comprise the following properties: i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 98% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format; and

vii) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC.

In certain embodiments,

i) the expression yield in Fab format was determined by UV-spectrophotometry using an extinction coefficient of 1.538 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

ii) the thermal stability in Fab format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

iii) the monomeric content (% monomer) in Fab format was determined by size exclusion chromatography using a Superdex75 HR10/30 column and Gibco D-PBS buffer at pH 7.4.

In certain embodiments,

iv) the expression yield in IgG1 format was determined by UV-spectrophotometry using an extinction coefficient of 1.369 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

v) the thermal stability in IgG1 format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

vi) the monomeric content (% monomer) in IgG1 format was determined by size exclusion chromatography using a Tosoh TSK-Gel G3000SWxl column and and Gibco D-PBS buffer at pH 7.4.

UV-spectrophotometry may be performed using the Nanadrop system (peqlab, Erlangen, Germany). Differential scanning fluorometry may be performed using the iCycler iQ5 Thermal Cycler (Biorad). Differential scanning fluorometry may be performed using Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). Size exclusion chromatography may be peformed using the ÄKTA Purifier System (GE Healthcare).

The following germline protein pairs (54) were at or above the following thresholds using the method described above: i) an expression yield in Fab format of at least 2.5 mg/L; ii) thermal stability at 70°C or above in Fab format; iii) monomeric content (% monomer) in Fab format of at

least 98% as determined by SEC; iv) an expression yield in IgG1 format of at least 30 mg/L; v) thermal stability at 73°C or above in IgG1 format; and vii) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC, therefore, have superior functional activity related to developability, (data shown in Figures 16-24): VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO:

239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252). Therefore, collections comprising any number of these germline protein pairs could be used to identify developable antibodies or fragments thereof against any antigen.

In an aspect, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of specific variable heavy chain and variable light chain pairs, for example, VH1-18/VK1-39. This means that the collection comprises antibodies or fragments wherein the framework regions of the antibodies or fragments comprise the germline protein sequences of VH1-18/VK1-39, where the variable heavy chain framework regions comprise the germline protein sequences of VH1-18 and the variable light chain framework regions comprise the germline protein sequences of VK1-39. A large number of germline protein pairs were tested, as constructs (as described in Examples 5 and 9), for their functional properties related to development. A number of constructs tested showed superior functional properties related to developability. The inventors believe that there is a high correlation between the input (antibody collection used for selection against an antigen) and output (antibodies identified from the collection as specific for the antigen) regarding the tested functional properties. Therefore, the collections of the invention comprise antibodies or fragments that comprise, in part, the same amino acid sequences as the constructs tested, for example, the framework regions and/or complementarity determining regions. Since, in an aspect, the collections comprise the amino acid sequences, or the nucleic acids encoding them, of the tested constructs it is believed that the collections comprise antibodies or fragments having the same superior functional properties related to developabiltiy as the constructs tested. Therefore, it is expected that the antibodies or fragments subsequently selected from the collections against an antigen will also have the same superior functional properties relevant to developability. This hypothesis is supported by the experiments and data described in Example 11, see Figures 37-39, 45-48 and 62.

In some embodiments, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences selected from two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, eleven

or more, twelve or more, thirteen or more, fourteen or more, fifiteen or more, sixteen or more, seventeen or more, eighteen or more, nineteen or more, twenty or more, twenty one or more, twenty two or more, twenty three or more, twenty four or more, twenty five or more, twenty six or more, twenty seven or more, twenty eight or more, twenty nine or more, thirty or more, thirty one or more, thirty two or more, thirty three or more, thirty four or more, thirty five or more, or thirty six or more, thirty seven or more, thirty eight or more, thirty nine or more, forty or more, forty one or more, or forty two or more, or forty three or more, or forty four or more, or forty five or more, or forty six or more, or forty seven or more, or forty eight or more, or forty nine or more, or fifty or more, or fifty one or more, or fifty two or more, or fifty three or more, or fifty four variable heavy chain and variable light chain pairs of VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06

(SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In an embodiment, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or at least 95% or each of the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences selected from the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15

(SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

An embodiment comprises a collection of synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of the, consisting of or consisting essentially of the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15

(SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In embodiments comprising the 54 pairs or a subset thereof, additional pairs may be selected to be added to the collection, wherein each germline protein pair added comprises the following properties:

i) an expression yield in Fab format of at least 2.5 mg/l as determined by UV-spectrophotometry using an extinction coefficient of 1.538 mL/mg and measuring absorbance at 280nm,

ii) thermal stability at 70°C or above in Fab format as determined by differential scanning fluorometry using PBS buffer,

iii) monomeric content (% monomer) in Fab format of at least 98% as determined by size exclusion chromatography using a Superdex75 HR10/30 column and Gibco D-PBS buffer at pH 7.4,

iv) an expression yield in IgG1 format of at least 30mg/l as determined by UV-spectrophotometry using an extinction coefficient of 1.369 mL/mg and measuring absorbance at 280nm,

v) thermal stability at 73°C or above in IgG1 format as determined by differential scanning fluorometry using PBS buffer, and

vi) the monomeric content (% monomer) in IgG1 format of at least 99% as determined by size exclusion chromatography using a Tosoh TSK-Gel G3000SWxl column and and Gibco D-PBS buffer at pH 7.4.

UV-spectrophotometry may be performed using the Nanadrop system (peqlab, Erlangen, Germany). Differential scanning fluorometry may be performed using the iCycler iQ5 Thermal Cycler (Biorad). Differential scanning fluorometry may be performed using Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). Size exclusion chromatography may be peformed using the ÄKTA Purifier System (GE Healthcare).

Embodiments of the present disclosure comprise subsets of the germline protein pairs (54) above having superior functional activity related to developability. In an embodiment, a subset of germline protein pairs (36 out of 54) were selected based upon a comparison of the stress testing data identified using the methods described in Examples 9.2.5 (a-d), data shown in Figures 19-24, Example 9.2.6 (a-d), data shown in Figures 49-54 and Example 9.2.7, scoring shown in Figures 55-60. The stress testing methods evaluated the 95 germline protein pairs in IgG1 format in order to determine their ability to withstand exposure to acid and agitation with glass beads. The 36 germline protein pairs, of an embodiment, were selected as they have additional superior functional properties relevant to developability as they showed strong resistance to acid and agitation stress. The 36 germline protein pairs selected in an embodiment, fulfilled all of the threshold functional activities of the 54, and, in addition, scored at or above 1225 in the stress testing cumulative score (as described in Example 9.2.7), which rated the germline protein pairs according to the following characteristics: absorption at 320 nm before and after acid exposure, radius and % polydispersity before and after acid exposure, particle staining before and after acid exposure, absorption at 320nm before and after agitation with glass beads, radius and % polydispersity after agitation with glass beads, and particle staining after agitation with glass beads. The 36 germline protein pairs selected in this embodiment, had values at or above the following thresholds for each criteria: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric content of purified Fab (as described in Example 9.1.3) of at least 98%; f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99% and g) stress testing cumulative score (as described in Example 9.2.7) of at least 1225.

Therefore, in an embodiment, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain

pairs comprise germline protein sequences of the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In embodiments comprising the 36 pairs or a subset thereof, additional pairs may be selected to be added to the collection, wherein each germline protein pair added comprises the following properties:

i) an expression yield in Fab format of at least 2.5 mg/l as determined by UV-spectrophotometry using an extinction coefficient of 1.538 mL/mg and measuring absorbance at 280nm,

ii) thermal stability at 70°C or above in Fab format as determined by differential scanning fluorometry using PBS buffer,

iii) monomeric content (% monomer) in Fab format of at least 98% as determined by size exclusion chromatography using a Superdex75 HR10/30 column and Gibco D-PBS buffer at pH 7.4,

iv) an expression yield in IgG1 format of at least 30mg/l as determined by UV-spectrophotometry using an extinction coefficient of 1.369 mL/mg and measuring absorbance at 280nm,

v) thermal stability at 73°C or above in IgG1 format as determined by differential scanning fluorometry using PBS buffer, and

vi) the monomeric content (% monomer) in IgG1 format of at least 99% as determined by size exclusion chromatography using a Tosoh TSK-Gel G3000SWxl column and and Gibco D-PBS buffer at pH 7.4.

UV-spectrophotometry may be performed using the Nanadrop system (peqlab, Erlangen, Germany). Differential scanning fluorometry may be performed using the iCycler iQ5 Thermal Cycler (Biorad). Differential scanning fluorometry may be performed using Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). Size exclusion chromatography may be peformed using the ÄKTA Purifier System (GE Healthcare).

In embodiments, a collection of synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or at least 95% or each of the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences selected from the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO:

213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In other embodiments, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences selected from two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, eleven or more, twelve or more, thirteen or more, fourteen or more, fifiteen or more, sixteen or more, seventeen or more, eighteen or more, nineteen or more, twenty or more, twenty one or more, twenty two or more, twenty three or more, twenty four or more, twenty five or more, twenty six or more, twenty seven or more, twenty eight or more, twenty nine or more, thirty or more, thirty one or more, thirty two or more, thirty three or more, thirty four or more, thirty five or more, or thirty six of the following variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO:

214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In an embodiment, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of, consisting of or consisting essentially of the following variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In another embodiment, the thresholds for each criterion were selected as follows: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric content of purified Fab (as described in Example 9.1.3) of at least 99%; f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99%; g) isoelectric point of purified IgG1 (as described in Example 9.2.4) of at least 8.3; and h) stress testing cumulative score (as described in Example 9.2.7) of at least 1225.

Therefore, in an embodiment, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain framework regions comprising germline protein sequences of a germline protein pair, wherein said germline protein pairs comprise the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 99% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format;

vi) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC, and

vii) an isoelectric point in IgG1 format of at least 8.3.

In certain embodiments,

i) the expression yield in Fab format was determined by UV-spectrophotometry using an extinction coefficient of 1.538 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

ii) the thermal stability in Fab format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

iii) the monomeric content (% monomer) in Fab format was determined by size exclusion chromatography using a Superdex75 HR10/30 column and Gibco D-PBS buffer at pH 7.4.

In certain embodiments,

iv) the expression yield in IgG1 format was determined by UV-spectrophotometry using an extinction coefficient of 1.369 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

v) the thermal stability in IgG1 format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

vi) the monomeric content (% monomer) in IgG1 format was determined by size exclusion chromatography using a Tosoh TSK-Gel G3000SWxl column and and Gibco D-PBS buffer at pH 7.4.

UV-spectrophotometry may be performed using the Nanadrop system (peqlab, Erlangen, Germany). Differential scanning fluorometry may be performed using the iCycler iQ5 Thermal Cycler (Biorad). Differential scanning fluorometry may be performed using Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). Size exclusion chromatography may be peformed using the ÄKTA Purifier System (GE Healthcare).

In additional embodiments, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein wherein substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90%, or at least 95% or each of the antibodies or functional fragments comprise variable heavy chain and variable light chain framework regions comprising germline protein sequences of a germline protein pair, wherein said germline protein pairs comprise the following properties: i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 99% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format;

vi) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC, and

vii) an isoelectric point in IgG1 format of at least 8.3.

In additional embodiments, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein wherein the antibodies or functional fragments consists of or consists essentially of variable heavy chain and variable light chain framework regions comprising germline protein

sequences of a germline protein pair, wherein said germline protein pairs comprise the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 99% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format;

vi) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC, and

vii) an isoelectric point in IgG1 format of at least 8.3.

The following germline protein pairs (33) were at or above the following thresholds: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric content of purified Fab (as described in Example 9.1.3) of at least 99%; f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99%; g) isoelectric point of purified IgG1 (as described in Example 9.2.4) of at least 8.3; and h) stress testing cumulative score (as described in Example 9.2.7) of at least 1225, therefore, have superior functional activity related to developability: VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO:

214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

Therefore, in an embodiment, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In embodiments comprising the 33 pairs or a subset thereof, additional pairs may be selected to be added to the collection, wherein each germline protein pair added comprises the following properties:

i) an expression yield in Fab format of at least 2.5 mg/l as determined by UV-spectrophotometry using an extinction coefficient of 1.538 mL/mg and measuring absorbance at 280nm,

ii) thermal stability at 70°C or above in Fab format as determined by differential scanning fluorometry using PBS buffer,

iii) monomeric content (% monomer) in Fab format of at least 99% as determined by size exclusion chromatography using a Superdex75 HR10/30 column and Gibco D-PBS buffer at pH 7.4,

iv) an expression yield in IgG1 format of at least 30mg/l as determined by UV-spectrophotometry using an extinction coefficient of 1.369 mL/mg and measuring absorbance at 280nm,

v) thermal stability at 73°C or above in IgG1 format as determined by differential scanning fluorometry using PBS buffer, and

vi) the monomeric content (% monomer) in IgG1 format of at least 99% as determined by size exclusion chromatography using a Tosoh TSK-Gel G3000SWxl column and and Gibco D-PBS buffer at pH 7.4.

UV-spectrophotometry may be performed using the Nanadrop system (peqlab, Erlangen, Germany). Differential scanning fluorometry may be performed using the iCycler iQ5 Thermal Cycler (Biorad). Differential scanning fluorometry may be performed using Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). Size exclusion chromatography may be peformed using the ÄKTA Purifier System (GE Healthcare).

In embodiments comprising the 33 pairs or a subset thereof, additional pairs may be selected to be added to the collection, wherein each germline protein pair added further comprises the following property:

vii) an isoelectric point in IgG1 format of at least 8.3.

In a further embodiment, pairs are added to a collection even though the pairs themselves did not meet all of the thresholds within each criteria, but were added to the collections in order to enhance diversity. In an embodiment the collection of 33 germline protein pairs further comprises: VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); and VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256). In this embodiment, the collection comprises (36 pairs): VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID

NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In embodiments, collections comprising any number of these germline protein pairs or synthetic nucleic acids encoding such antibodies or functional fragments could be used to identify developable antibodies or fragments thereof against any antigen.

In some embodiments, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences selected from two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, eleven or more, twelve or more, thirteen or more, fourteen or more, fifiteen or more, sixteen or more, seventeen or more, eighteen or more, nineteen or more, twenty or more, twenty one or more, twenty two or more, twenty three or more, twenty four or more, twenty five or more, twenty six or more, twenty seven or more, twenty eight or more, twenty nine or more, thirty or more, thirty one or more, thirty two or more, thirty three or more variable heavy chain and variable light chain pairs selected from the group consisting of VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51

(SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In an embodiment, a collection comprises synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or at least 95% or each of the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences selected from the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27

(SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

An embodiment comprises a collection of synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, wherein the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of, consisting of or consisting essentially of the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09

(SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

An additional aspect to the present invention is the ability of the collections to be useful in identifying antibodies or functional fragments thereof against any immunogen. Therefore, in some embodiments, the collections comprise variable heavy chain framework regions and variable light chain framework regions comprising germline protein sequences of at least two different germline protein pairs; at least three different germline protein pairs; at least four different germline protein pairs; at least five different germline protein pairs; at least six different germline protein pairs; at least seven different germline protein pairs; at least eight different germline protein pairs; at least nine different germline protein pairs; at least ten different germline protein pairs; at least eleven different germline protein pairs; at least twelve different germline protein pairs; at least thirteen different germline protein pairs; at least fourteen different germline protein pairs; at least fifteen different germline protein pairs; at least sixteen different germline protein pairs; at least seventeen different germline protein pairs; at least eighteen different germline protein pairs; at least nineteen different germline protein pairs; at least twenty different germline protein pairs; at least 21 different germline protein pairs; at least 22 different germline protein pairs; at least 23 different germline protein pairs; at least 24 different germline protein pairs; at least 25 different germline protein pairs; at least 26 different germline protein pairs; at least 27 different germline protein pairs; at least 28 different variable heavy chain germline protein; at least 29 different germline protein pairs sequences; at least 30 different germline protein pairs; at least 31 different germline protein pairs; at least 32 different germline protein pairs; at least 33 different germline protein pairs; at least 34 different germline protein pairs; at least 35 different germline protein pairs; at least 36 different germline protein pairs; at least 37 different germline protein pairs; at least 38 different germline protein pairs; at least 39 different germline protein pairs; at least 40 different germline protein pairs; at least 41 different germline protein pairs; at least 42 different germline protein pairs; at least 43 different germline protein pairs; at least 44 different germline protein pairs; at least 45 different germline protein pairs; at least 46 different germline protein pairs; at least 47 different germline protein pairs; at least 48 different germline protein pairs; at least 49 different germline protein pairs; at least 50 different germline protein pairs; at least 51 different germline protein pairs; at least 52 different germline protein pairs; at least 53 different germline protein pairs; at least 54 different germline protein pairs.

As a low potential for immunogenicity in humans is a goal for therapeutic antibodies, in an aspect, the collections comprise framework regions comprising germline protein sequences

or nucleic acids encoding them. In addition, in order to maintain a low risk of immunogenicity, complementarity determining regions may be used comprising germline protein sequences. In an embodiment, the collections comprise synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising one or more complementarity determining regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the complementarity determining regions of the variable heavy chains and variable light chains are depicted in Figures 25-33. More specifically, in an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising CDR1 regions comprising germline protein sequences from the respective variable heavy chain and/or variable light chain pairs, wherein the amino acid and nucleic acid sequences of the CDR1 region of the variable heavy chains and variable light chains are depicted in Figures 25, 28, and 31, and the corresponding SEQ ID NOs: 204-265. In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR1 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the HCDR1 region of the variable heavy chains and variable light chains are depicted in Figure 25 and the corresponding SEQ ID NOs: 204-229. In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising LCDR1 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the LCDR1 region of the variable heavy chains and variable light chains are depicted in Figures 28 and 31 and the corresponding SEQ ID NOs: 230-265. In an additional embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising CDR2 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the CDR2 region of the variable heavy chains and variable light chains are depicted in Figures 26, 29, and 32, and the corresponding SEQ ID NOs: 204-265. In an additional embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR2 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the HCDR2

region of the variable heavy chains and variable light chains are depicted in Figure 26 and the corresponding SEQ ID NOs: 204-229. In an additional embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising LCDR2 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino and nucleic acid sequences of the LCDR2 region of the variable heavy chains and variable light chains are depicted in Figures 29 and 32 and the corresponding SEQ ID NOs: 230-265.

An aspect of the disclosure includes modifying germline complementarity determining regions to remove potential post translational modification sites (PTMs). Examples of variable heavy chain complementarity determining regions modified to remove PTMs are shown in Figures 34-36. In an aspect, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising one or more complementarity determining regions comprising amino acid modifications that remove potential post translational modification sites. In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising one or more complementarity determining regions comprising the complementarity determing region sequences or nucleic acid sequences encoding the same depicted in Figures 34-36 from the respective variable heavy chain. In a further embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR1 regions comprising the HCDR1 or nucleic acids encoding the same depicted in Figures 34-36 from the respective variable heavy chain. The amino acid sequences of the low PTM HCDR1s are depicted in SEQ ID NOs: 266-278. The nucleic acid sequences of the low PTM HCDR1s are depicted in SEQ ID NOs: 279-291. In a further embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR2 regions comprising the HCDR2 regions or nucleic acids encoding the same depicted in Figures 34-36 from the respective variable heavy chain. The amino acid sequences of the low PTM HCDR2s are depicted in SEQ ID NOs: 266-278. The nucleic acid sequences of the low PTM HCDR2s are depicted in SEQ ID NOs: 279-291.

An aspect of the disclosure includes utilizing germline FR4 sequences in the collections. In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a FR4 region selected from the group consisting of: JH4 (SEQ ID NO:293), Jκ1 (SEQ ID NO:297), and Jλ2/3 (SEQ ID NO:301). In an embodiment, a collection comprises antibodies or functional

fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a germline JH4 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The JH4 FR4 amino acid sequence is depicted in (SEQ ID NO:293) and (SEQ ID NO:295). The JH4 FR4 nucleic acid sequence is depicted in (SEQ ID NO:292) and (SEQ ID NO:294). In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a germline Jk1 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The Jk1 FR4 amino acid sequence is depicted in (SEQ ID NO:297). The Jk1 FR4 nucleic acid sequence is depicted in (SEQ ID NO:296), (SEQ ID NO:298) and (SEQ ID NO:299). In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a germline Jλ2/3 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The Jλ2/3 FR4 amino acid sequence is depicted in (SEQ ID NO:301). The Jλ2/3 FR4 nucleic acid sequence is depicted in (SEQ ID NO:300), (SEQ ID NO:302) and (SEQ ID NO:303).

In an aspect, in order to enhance the ability of identifying antibodies or fragment thereof against any antigen, collections comprise a diversified CDR3 region. In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a diversified HCDR3 region. In an embodiment, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a diversified LCDR3 region.

In another aspect, in order to enhance the ability of identifying antibodies or fragments thereof against any antigen, collections comprise at least 1×10^4 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^5 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^6 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^7 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^8 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^9 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^{10} antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, or at least 1×10^{11} antibodies

or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments.

In an embodiment the collections comprise antibodies or synthetic nucleic acids encoding such antibodies selected from the group consisting of human IgG1, IgG2, IgG3, IgG4, IgA, IgE, IgM and IgD. In an embodiment the collections comprise antibody fragments or synthetic nucleic acids encoding such fragments selected from the group consisting of Fab, F(ab')2, Fab', Fv, and scFv.

In embodiments, the IgG heavy chain constant domains of the antibodies of the collections comprise the amino acid sequences shown in Figures 41A-B (SEQ ID NO: 305). In other embodiments, the nucleic acids encoding the IgG heavy chain constant domains of the antibodies of the collection comprise the nucleic acid sequences shown in Figures 41A-B (SEQ ID NO: 304). In embodiments, the Fab heavy chain constant domains of the antibody fragments of the collections comprise the amino acid sequences shown in Figure 42 (SEQ ID NO: 307). In other embodiments, the nucleic acids encoding the Fab heavy chain constant domains of the antibodies of the collection comprise the nucleic acid sequences shown in Figure 42 (SEQ ID NO: 306). In embodiments, the IgG (SEQ ID NO: 309) and/or Fab (SEQ ID NO: 311) kappa light chain constant domains of the antibodies or antibody fragments of the collections comprise the amino acid sequences shown in Figure 43. In other embodiments, the nucleic acids encoding the IgG (SEQ ID NO: 308) and/or Fab (SEQ ID NO: 310) kappa light chain constant domains of the antibodies or antibody fragments of the collections comprise the nucleic acid sequences shown in Figure 43. In embodiments, the IgG (SEQ ID NO: 313) and/or Fab (SEQ ID NO: 315) lambda light chain constant domains of the antibodies or antibody fragments of the collections comprise the amino acid sequences shown in Figure 44. In other embodiments, the nucleic acids encoding the IgG (SEQ ID NO: 312) and/or Fab (SEQ ID NO: 314) lambda light chain constant domains of the antibodies or antibody fragments of the collections comprise the nucleic acid sequences shown in Figure 44.

An aspect comprises, a vector comprising the collections of nucleic acids described herein. In an embodiment, the vector comprises a display vector. In an embodiment, the vector comprises a phagemid vector, yeast display or mammalian display vector. An aspect is a recombinant host cell comprising the nucleic acids described herein, or a vector described herein. In an embodiment, the recombinant host is prokaryotic or eukaryotic. In embodiment, the recombinant host cell of is *E. coli*, mammalian or yeast.

Methods of Making

An aspect comprises methods of producing the collections described herein.

An aspect comprises, a method of producing a collection of synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, comprising

a) identifying the variable heavy chain and variable light chain germline gene pairs present in the human immune repertoire;

b) testing the variable heavy chain and variable light chain germline protein pairs identified in step a) for the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 98% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format; and

vii) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC; and

c) generating a collection, wherein substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or at least 95% or each of the, or the antibodies or functional fragments thereof comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of the germline protein pairs fulfilling the properties of step b).

In certain embodiments of the method,

i) the expression yield in Fab format was determined by UV-spectrophotometry using an extinction coefficient of 1.538 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

ii) the thermal stability in Fab format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

iii) the monomeric content (% monomer) in Fab format was determined by size exclusion chromatography using a Superdex75 HR10/30 column and Gibco D-PBS buffer at pH 7.4.

In certain embodiments,

iv) the expression yield in IgG1 format was determined by UV-spectrophotometry using an extinction coefficient of 1.369 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

v) the thermal stability in IgG1 format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

vi) the monomeric content (% monomer) in IgG1 format was determined by size exclusion chromatography using a Tosoh TSK-Gel G3000SWxl column and and Gibco D-PBS buffer at pH 7.4.

UV-spectrophotometry may be performed using the Nanadrop system (peqlab, Erlangen, Germany). Differential scanning fluorometry may be performed using the iCycler iQ5 Thermal Cycler (Biorad). Differential scanning fluorometry may be performed using Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). Size exclusion chromatography may be peformed using the ÄKTA Purifier System (GE Healthcare).

In an embodiment, step a) further comprises the steps

aa) isolating human B cells from a sample;

ab) generating cDNAs from the B cells;

ac) PCR amplifying the cDNAs from the B cells;

ad) sequencing the PCR products;

ae) identifying the germline genes of each PCR product.

The DNA encoding antibodies and fragments thereof from each B cell are isolated, and amplified e.g., the heavy and light chain are physically isolated in a PCR reaction. The DNA is preferably sequenced. The DNA sequenced may be cDNA generated from B cell mRNA. mRNA extraction from eukaryotic cells, such as B cells, is a well know technological procedure. Numerous protocols exist and commercial kits are available. Such as the PolyATtract® mRNA Isolation System (Promega, Madison, WI, USA) or various RNeasy and Oligotex DirectmRNA kits (both from Qiagen, Hilden, Germany). Many of these techniques make use of the polyA tail of the eukaryotic mRNA, e.g. via affinity purification to oligo (dT) matrices, such as oligo (dT) cellulose.

cDNA can be selectively amplified from the isolated mRNA via reverse transcription using specific primers, followed by conventional PCR. Specific primers are used to amplify variable heavy and light chain domain nucleic acids. See Cancer Surv. 1997;30:21–44, J Clin. Pathol. 1994;47:493–6, J. Clin. Pathol. 1990;43:888–90 or Mol. Pathol. 2002 April; 55(2): 98–101. The DNA coding for both the variable and light chain domains from one B cell are

maintained together so that the variable domain heavy and light chain class pairing can be identified. Techniques for the isolation of nucleic acids encoding variable domain pairings from individual B cells are well known in the art. See for example, WO01/92291; WO92/15678; WO93/03151, WO2005/042774; Mullinax RL et al., 1992 Biotechniques 12:6 864-868; Chapal, N. et al. 1997 Biotechniques 23, 518-524, , Embleton MJ et al., 1992 Nucleic Acids Res. 20:15, 3831-3837; Coronella, J.A. et al. 2000 Nucleic Acids Res. 28:20, E85; Thirion S et al., 1996 European Journal of Cancer Prevention 5:6 507-511; and Wang, X et al. 2000 J. Immunol. Methods 20, 217-225.

Preferably, the DNA from each of the B cells is sequenced. Various companies exist which are able to sequence entire genomes, such as Helicos BioSciences Corporation (Cambridge, MA, USA). With its True Single Molecule Sequencing™ technology, Helicos is able to directly sequence single molecules of DNA or RNA at high speed and efficiency. Other companies able to perform similar sequence endeavors include Illumina (San Diego, CA, USA; Solexa system) and Roche (Basel, CH; 454 system). No cloning steps are required prior to sequencing.

In another aspect, the disclosure enables methods of identifying the germline family of the heavy and light chain variable domain pairs present in the immune repertoire. All antibodies or fragments thereof can be traced back to their germline family using methods known to one of skill in the art. By analyzing the sequence of a nucleic acid encoding an antibody or fragment thereof, the germline family of both the VH and VL can be determined by methods known to one of skill in the art. For example, Wildt et. al, (1999) sampled B cells from 3 patients and identified 365 VH and VL class pairings. The RNA from each B cell was used for cDNA synthesis and the cDNA encoding the VH and VL regions was PCR amplified and sequenced. As shown in Fig. 1 of Wildt, certain VH and VLs classes paired more frequently than others, for example, VH3-8 with Vκ3-1, Vκ3-19, Vκ4-1, Vλ2-3, or Vλ1-2, and VH3-9 with Vκ3-1, Vκ3-3 or Vλ1-5.

In an embodiment, step b) further comprises the steps

ba) synthesizing DNA encoding antibodies or functional fragments thereof comprising variable heavy chain and variable light chain germline protein pairs representing the pairs present in the human immune repertoire;

bb) expressing the germline protein pairs synthesized in ba); and

bc) testing the germline protein pairs of bb) for each of the properties.

In an aspect of the method, the nucleic acids encoding collections of antibodies or fragments thereof of the invention are synthesized and expressed in collections that may be used for selection against an antigen. In this embodiment the method comprises step c),

wherein step c) comprises the steps ca) synthesizing nucleic acids encoding the antibodies or functional fragments thereof; cb) cloning the nucleic acids into a vector; cc) expressing the antibodies or functional fragments thereof.

In another embodiment of the method, the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise the germline protein sequences from the variable heavy chain and variable light chain pairs of VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In another embodiment of the method, the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences comprising two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, eleven or more, twelve or more, thirteen or more, fourteen or more, fifiteen or more, sixteen or more, seventeen or more, eighteen or more,

nineteen or more, twenty or more, twenty one or more, twenty two or more, twenty three or more, twenty four or more, twenty five or more, twenty six or more, twenty seven or more, twenty eight or more, twenty nine or more, thirty or more, thirty one or more, thirty two or more, thirty three or more, thirty four or more, thirty five or more, or thirty six or more, thirty seven or more, thirty eight or more, thirty nine or more, forty or more, forty one or more, or forty two or more, or forty three or more, or forty four or more, or forty five or more, or forty six or more, or forty seven or more, or forty eight or more, or forty nine or more, or fifty or more, or fifty one or more, or fifty two or more, or fifty three or more, or fifty four variable heavy chain and variable light chain pairs selected from the group consisting of VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74

(SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In further embodiments of the method, substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or at least 95% or each of the, or the antibodies or functional fragments comprises variable heavy chain and variable light chain framework regions comprising germline protein sequences of a germline protein pair, wherein said germline protein pair comprises the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 99% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format; and

vii) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC

viii) an isoelectric point in IgG1 format of at least 8.3.

In this embodiment of the method, the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise the germline protein sequences from the variable heavy chain and variable light chain pairs of VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO:

237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In a further embodiment of the method, the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs further comprise the germline protein sequences from the variable heavy chain and variable light chain pairs of VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); and VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256). In this embodiment of the method, the collection comprises (36 pairs): VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO:

215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

Methods of Using

An aspect comprises methods of using the collections described herein to identify antibodies or fragments specific for an antigen.

An aspect of the disclosure comprises a method of identifying an antibody or antibody fragment specific for an antigen, comprising:

(a) contacting the antigen with a collection of antibodies or functional fragments thereof, wherein substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or at least 95% or each of the, or the antibodies or functional fragments of the collection comprise variable heavy chain and variable light chain pairs, wherein the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences of germline protein pairs comprising the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 98% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format; and

vii) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC, and

(b) selecting one or more antibodies or antibody fragments that bind to said antigen.

In certain embodiments,

i) the expression yield in Fab format was determined by UV-spectrophotometry using an extinction coefficient of 1.538 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

ii) the thermal stability in Fab format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

iii) the monomeric content (% monomer) in Fab format was determined by size exclusion chromatography using a Superdex75 HR10/30 column and Gibco D-PBS buffer at pH 7.4.

In certain embodiments,

iv) the expression yield in IgG1 format was determined by UV-spectrophotometry using an extinction coefficient of 1.369 mL/mg and measuring absorbance at 280nm.

In certain embodiments,

v) the thermal stability in IgG1 format was determined by differential scanning fluorometry using PBS buffer.

In certain embodiments,

vi) the monomeric content (% monomer) in IgG1 format was determined by size exclusion chromatography using a Tosoh TSK-Gel G3000SWxl column and and Gibco D-PBS buffer at pH 7.4.

UV-spectrophotometry may be performed using the Nanadrop system (peqlab, Erlangen, Germany). Differential scanning fluorometry may be performed using the iCycler iQ5 Thermal Cycler (Biorad). Differential scanning fluorometry may be performed using Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). Size exclusion chromatography may be peformed using the ÄKTA Purifier System (GE Healthcare).

In an embodiment of the method the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences from VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO:

239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In an embodiment of the method the antibodies or functional fragments thereof comprise two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, eleven or more, twelve or more, thirteen or more, fourteen or more, fifiteen or more, sixteen or more, seventeen or more, eighteen or more, nineteen or more, twenty or more, twenty one or more, twenty two or more, twenty three or more, twenty four or more, twenty five or more, twenty six or more, twenty seven or more, twenty eight or more, twenty nine or more, thirty or more, thirty one or more, thirty two or more, thirty three or more, thirty four or more, thirty five or more, or thirty six or more, thirty seven or more, thirty eight or more, thirty nine or more, forty or more, forty one or more, or forty two or more, or forty three or more, or forty four or more, or forty five or more, or forty six or more, or forty seven or more, or forty eight or more, or forty nine or more, or fifty or more, or fifty one or more, or fifty two or more, or fifty three or more, or fifty four variable heavy chain and variable light chain pairs selected from the group consisting of variable heavy chain and variable light chain pairs selected from VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15

(SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In embodiment of the method, substantially all, or at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or at least 95% or each of the, or the antibodies or functional fragments comprises variable heavy chain and variable light chain framework regions comprising germline protein sequences of a germline protein pair, wherein said germline protein pair comprises the following properties:

i) an expression yield in Fab format of at least 2.5 mg/L;

ii) thermal stability at 70°C or above in Fab format;

iii) monomeric content (% monomer) in Fab format of at least 99% as determined by SEC;

iv) an expression yield in IgG1 format of at least 30 mg/L;

v) thermal stability at 73°C or above in IgG1 format;

vii) monomeric content (% monomer) in IgG1 format of at least 99% as determined by SEC; and

viii) an isoelectric point in IgG1 format of at least 8.3.

In this embodiment of the method the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences from VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO:

207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In a further embodiment of the method the framework regions of the variable heavy chain and variable light chain pairs comprise germline protein sequences further selected from the variable heavy chain and variable light chain pairs VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); and VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256). In this embodiment, a collection comprises (36 pairs): VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23

(SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

Aspects of the Methods

In further aspects of the methods disclosed herein, the collections comprise synthetic antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising one or more complementarity determining regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid sequences of the complementarity determining regions of the variable heavy chains and variable light chains are depicted in Figures 25-33. More specifically, in an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising CDR1 regions comprising germline protein sequences from the respective variable heavy chain and/or variable light chain pairs, wherein the amino acid and nucleic acid sequences of the CDR1 region of the variable heavy chains and variable light chains are depicted in Figures 25, 28 and 31 and the corresponding SEQ ID NOs: 204-265. In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR1 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the HCDR1 region of the variable heavy chains and variable light chains are depicted in Figures 25 and the corresponding SEQ ID NOs: 204-229. In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising LCDR1 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the LCDR1 region of the variable heavy chains and variable light chains are depicted in Figures 28 and 31 and the corresponding SEQ ID NOs: 230-265. In an

additional embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising CDR2 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the CDR2 region of the variable heavy chains and variable light chains are depicted in Figures 26, 29 and 32 and the corresponding SEQ ID NOs:204-265. In an additional embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR2 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the HCDR2 region of the variable heavy chains and variable light chains are depicted in Figures 26 and the corresponding SEQ ID NOs: 204-229. In an additional embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising LCDR2 regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pairs, wherein the amino acid and nucleic acid sequences of the LCDR2 region of the variable heavy chains and variable light chains are depicted in Figures 29 and 32 and the corresponding SEQ ID NOs: 230-265.

In embodiments of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising one or more complementarity determining regions comprising amino acid modifications that remove potential post translational modification sites. In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising one or more complementarity determining regions comprising the complementarity determing region sequences or nucleic acid sequences encoding the same depicted in Figures 34-36 from the respective variable heavy chain. In a further embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR1 regions comprising the HCDR1 or nucleic acids encoding the same depicted in Figures 34-36 from the respective variable heavy chain. The amino acid sequences of the low PTM HCDR1s are depicted in SEQ ID NOs: 266-278. The nucleic acid sequences of the low PTM HCDR1s are depicted in SEQ ID NOs: 279-291. In a further embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising HCDR2 regions comprising the HCDR2 regions or nucleic acids encoding the same depicted in Figures

34-36 from the respective variable heavy chain. The amino acid sequences of the low PTM HCDR2s are depicted in SEQ ID NOs: 266-278. The nucleic acid sequences of the low PTM HCDR2s are depicted in SEQ ID NOs: 279-291.

In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a FR4 region selected from the group consisting of: JH4 (SEQ ID NO:293), Jκ1 (SEQ ID NO:297), and Jλ2/3 (SEQ ID NO:301). In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a germline JH4 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The JH4 FR4 amino acid sequence is depicted in (SEQ ID NO:293) and (SEQ ID NO:295). The JH4 FR4 nucleic acid sequence is depicted in (SEQ ID NO:292) and (SEQ ID NO:294). In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a germline Jk1 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The Jk1 FR4 amino acid sequence is depicted in (SEQ ID NO:297). The Jk1 FR4 nucleic acid sequence is depicted in (SEQ ID NO:296), (SEQ ID NO:298) and (SEQ ID NO:299). In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a germline Jλ2/3 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The Jλ2/3 FR4 amino acid sequence is depicted in (SEQ ID NO:301). The Jλ2/3 FR4 nucleic acid sequence is depicted in (SEQ ID NO:300), (SEQ ID NO:302) and (SEQ ID NO:303).

In an aspect, in order to enhance the ability of identifying antibodies or fragment thereof against any antigen, collections comprise a diversified CDR3 region. In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a diversified HCDR3 region. In an embodiment of the method, a collection comprises antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments comprising a diversified LCDR3 region.

In another aspect, in order to enhance the ability of identifying antibodies or fragments thereof against any antigen, collections of the method comprise at least 1×10^4 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^5 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^6 antibodies or functional

fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^7 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^8 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^9 antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, at least 1×10^{10} antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments, or at least 1×10^{11} antibodies or functional fragments thereof or synthetic nucleic acids encoding such antibodies or functional fragments.

In an embodiment of the method the collections comprise antibodies or synthetic nucleic acids encoding such antibodies selected from the group consisting of human IgG1, IgG2, IgG3, IgG4, IgA, IgE, IgM and IgD. In an embodiment of the method the collections comprise antibody fragments or synthetic nucleic acids encoding such fragments selected from the group consisting of Fab, F(ab')2, Fab', Fv, and scFv.

In embodiments of the method, the IgG heavy chain constant domains of the antibodies of the collections comprise the amino acid sequences shown in Figures 41A-B (SEQ ID NO: 305). In other embodiments of the method, the nucleic acids encoding the IgG heavy chain constant domains of the antibodies of the collection comprise the nucleic acid sequences shown in Figures 41A-B (SEQ ID NO: 304). In embodiments, the Fab heavy chain constant domains of the antibody fragments of the collections comprise the amino acid sequences shown in Figure 42 (SEQ ID NO: 307). In other embodiments of the method, the nucleic acids encoding the Fab heavy chain constant domains of the antibodies of the collection comprise the nucleic acid sequences shown in Figure 42 (SEQ ID NO: 306). In embodiments of the method, the IgG (SEQ ID NO: 309) and/or Fab (SEQ ID NO: 311) kappa light chain constant domains of the antibodies or antibody fragments of the collections comprise the amino acid sequences shown in Figure 43. In other embodiments of the method, the nucleic acids encoding the IgG (SEQ ID NO: 308) and/or Fab (SEQ ID NO: 310) kappa light chain constant domains of the antibodies or antibody fragments of the collections comprise the nucleic acid sequences shown in Figure 43. In embodiments of the method, the IgG (SEQ ID NO: 313) and/or Fab (SEQ ID NO: 315) lambda light chain constant domains of the antibodies or antibody fragments of the collections comprise the amino acid sequences shown in Figure 44. In other embodiments of the method, the nucleic acids encoding the IgG (SEQ ID NO: 312) and/or Fab (SEQ ID NO: 314) lambda light chain constant domains of the antibodies or antibody fragments of the collections comprise the nucleic acid sequences shown in Figure 44.

Antibodies of the Invention

In another aspect, the disclosure provides a synthetic antibody or functional fragment thereof or a synthetic nucleic acid encoding such antibody or functional fragments, wherein the antibody or functional fragment comprises a variable heavy chain and variable light chain pair, wherein the framework regions of the variable heavy chain and variable light chain pair comprises germline protein sequences selected from the variable heavy chain and variable light chain pairs VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51

(SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises one or more complementarity determining regions comprising germline protein sequences from the respective variable heavy chain and variable light chain pair, wherein the amino acid sequence of the complementarity determining region of the variable heavy chain and variable light chain are depicted in Figures 25-33. More specifically, in an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises a CDR1 region comprising germline protein sequences from the respective variable heavy chain and/or variable light chain pair, wherein the amino acid sequence of the CDR1 region of the variable heavy chains and variable light chains are depicted in Figures 25, 28, and 31, and the corresponding SEQ ID NOs: 204-265. In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises an HCDR1 region comprising germline protein sequences from the respective variable heavy chain and variable light chain pair, wherein the amino acid sequences of the HCDR1 region of the variable heavy chain and variable light chain are depicted in Figure 25 and the corresponding SEQ ID NOs: 204-229. In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises an LCDR1 region comprising germline protein sequences from the respective variable heavy chain and variable light chain pair, wherein the amino acid sequences of the LCDR1 region of the variable heavy chains and variable light chain are depicted in Figures 28 and 31 and the corresponding SEQ ID NOs: 230-265. In an additional embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises a CDR2 region comprising germline protein sequences from the respective variable heavy chain and variable light chain pair, wherein the amino acid sequences of the CDR2 region of the variable heavy chain and variable light chain are depicted in Figures 26, 29, and 32, and the corresponding SEQ ID NOs: 204-265. In an additional embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises an HCDR2 region comprising germline protein sequences from the

respective variable heavy chain and variable light chain pair, wherein the amino acid sequences of the HCDR2 region of the variable heavy chain and variable light chain are depicted in Figure 26 and the corresponding SEQ ID NOs: 204-229. In an additional embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises an LCDR2 region comprising germline protein sequences from the respective variable heavy chain and variable light chain pair, wherein the amino acid sequences of the LCDR2 region of the variable heavy chain and variable light chain are depicted in Figures 29 and 32 and the corresponding SEQ ID NOs: 230-265.

An aspect of the disclosure includes modifying germline complementarity determining regions to remove potential post translational modification sites (PTMs). Examples of variable heavy chain complementarity determining regions modified to remove PTMs are shown in Figures 34-36. In an aspect, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises one or more complementarity determining regions comprising amino acid modifications that remove potential post translational modification sites. In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises one or more complementarity determining regions comprising the complementarity determing region sequences or nucleic acid sequences encoding the same depicted in Figures 34-36 from the respective variable heavy chain. In a further embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises an HCDR1 region comprising the HCDR1 or nucleic acids encoding the same depicted in Figures 34-36 from the respective variable heavy chain. The amino acid sequences of the low PTM HCDR1s are depicted in SEQ ID NOs: 266-278. The nucleic acid sequences of the low PTM HCDR1s are depicted in SEQ ID NOs: 279-291. In a further embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises an HCDR2 region comprising the HCDR2 region or nucleic acids encoding the same depicted in Figures 34-36 from the respective variable heavy chain. The amino acid sequences of the low PTM HCDR2s are depicted in SEQ ID NOs: 266-278. The nucleic acid sequences of the low PTM HCDR2s are depicted in SEQ ID NOs: 279-291.

An aspect of the disclosure includes utilizing germline FR4 sequences. In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises a FR4 region selected from the group consisting of: JH4 (SEQ ID NO:293), Jκ1 (SEQ ID NO:297), and Jλ2/3 (SEQ ID

NO:301). In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises a germline JH4 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The JH4 FR4 amino acid sequence is depicted in (SEQ ID NO:293) and (SEQ ID NO:295). The JH4 FR4 nucleic acid sequence is depicted in (SEQ ID NO:292) and (SEQ ID NO:294). In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises a germline Jk1 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The Jk1 FR4 amino acid sequence is depicted in (SEQ ID NO:297). The Jk1 FR4 nucleic acid sequence is depicted in (SEQ ID NO:296), (SEQ ID NO:298) and (SEQ ID NO:299). In an embodiment, the synthetic antibody or functional fragment thereof or synthetic nucleic acid encoding such antibody or functional fragment thereof comprises a germline Jλ2/3 FR4 region, whose amino acid or nucleic acid sequence is depicted in Figure 40. The Jλ2/3 FR4 amino acid sequence is depicted in (SEQ ID NO:301). The Jλ2/3 FR4 nucleic acid sequence is depicted in (SEQ ID NO:300), (SEQ ID NO:302) and (SEQ ID NO:303).

In an embodiment the the synthetic antibody or synthetic nucleic acid encoding such antibody is selected from the group consisting of human IgG1, IgG2, IgG3, IgG4, IgA, IgE, IgM and IgD. In an embodiment the synthetic antibody fragment or synthetic nucleic acid encoding such antibody fragment is selected from the group consisting of Fab, F(ab')2, Fab', Fv, and scFv.

In embodiments, the IgG heavy chain constant domain of the antibody comprises the amino acid sequences shown in Figures 41A-B (SEQ ID NO: 305). In other embodiments, the nucleic acids encoding the IgG heavy chain constant domains of the antibody comprises the nucleic acid sequences shown in Figures 41A-B (SEQ ID NO: 304). In embodiments, the Fab heavy chain constant domain of the antibody fragments comprises the amino acid sequences shown in Figure 42 (SEQ ID NO: 307). In other embodiments, the nucleic acids encoding the Fab heavy chain constant domain of the antibody fragment comprises the nucleic acid sequences shown in Figure 42 (SEQ ID NO: 306). In embodiments, the IgG (SEQ ID NO: 309) and/or Fab (SEQ ID NO: 311) kappa light chain constant domains of the antibodies or antibody fragments comprise the amino acid sequences shown in Figure 43. In other embodiments, the nucleic acids encoding the IgG (SEQ ID NO: 308) and/or Fab (SEQ ID NO: 310) kappa light chain constant domains of the antibodies or antibody fragments comprise the nucleic acid sequences shown in Figure 43. In embodiments, the IgG (SEQ ID NO: 313) and/or Fab (SEQ ID NO: 315) lambda light chain constant domains of the antibodies or antibody fragments

comprise the amino acid sequences shown in Figure 44. In other embodiments, the nucleic acids encoding the IgG (SEQ ID NO: 312) and/or Fab (SEQ ID NO: 314) lambda light chain constant domains of the antibodies or antibody fragments comprise the nucleic acid sequences shown in Figure 44.

EXAMPLES

Example 1: Generation of restriction sites in the C-terminus of a prokaryotic signal sequence and human leader sequence, providing for fully germline FR1 regions

In one aspect, the present disclosure describes collections of antibodies or fragments thereof comprising framework regions comprising germline protein sequences, specifically FR1. It is expected that having germline sequences shall lower the immunogenicity risk of the antibodies when administered in humans. Compatible restriction sites, however, must be used in order to enable standard cloning of the nucleic acids encoding the collections of antibodies into display and/or expression vectors so that the antibodies can be screened against immunogens. In the past, restriction sites utilized for cloning were often located within the framework regions, thus modifying the nucleic acid and/or amino acid sequence away from germline. In order to ensure that at least the framework 1 (FR1) region of each of the antibodies of the present disclosure maintain a germline protein sequence, there should not be any restriction sites within FR1 which would lead to deviations from the germline amino acid sequence. Therefore, an aspect of the present disclosure is the incorporation of an identical or at least compatible restriction site within the C-terminus of prokaryotic signal sequences and human leader sequences, specifically within the three C-terminal residues. Additionally, a prokaryotic signal sequence and human leader sequence comprising an identical or compatible restriction site must be functional and allow for good display and expression yield of the antibodies or fragments thereof in both prokaryotic and mammalian expression systems.

Fig. 1 shows the selected restriction sites and their corresponding positions. The NheI (VLA) restriction site was selected for incorporation into the prokaryotic heavy chain signal sequences (phoA). The nucleic acid and amino acid sequences of the wildtype phoA signal sequence and the NheI (VLA) phoA signal sequence are shown in Table 1.

Table 1:

Wildtype E. coli phoA signal sequence (C-terminal amino acid sequence from position -3 to -1 is TKA without restriction site):

M K Q S T I A L A L L P L L F T P V T K A

ATGAAACAGAGCACCATTGCCCTGGCCCTGCTGCCGCTGCTGTTTACCCCAGTGACCAAA GCC

(SEQ ID NOS 8 and 7, respectively, in order of appearance)

PhoA wild type C-terminus	T	K	A
	ACC	AAA	GCC

Modified E. coli phoA signal sequence with C-terminal VLA and NheI restriction site (= GCTAGC):

M K Q S T I A L A L L P L L F T P V V L A

ATGAAACAGAGCACCATTGCCCTGGCCCTGCTGCCGCTGCTGTTTACCCCAGTGGT**GCTA GC**C

(SEQ ID NOS 10 and 9, respectively, in order of appearance)

The NdeI (AYA) restriction site was selected for incorporation into the prokaryotic kappa and lambda signal sequences (ompA). The nucleic acid and amino acid sequences of the wildtype ompA signal sequence and the modified NdeI (AYA) ompA signal sequence are shown in Table 2.

Table 2:

Wildtype E. coli ompA signal sequence (C-terminal amino acid sequence from position -3 to -1 is AQA without restriction site):

M K K T A I A I A V A L A G F A T V A Q A

ATGAAAAAAACCGCCATTGCCATTGCCGTGGCCCTGGCAGGCTTTGCCACCGTGGCGCAG GCC

(SEQ ID NOS 12 and 11, respectively, in order of appearance)

OmpA wild type C-terminus	A	Q	A
	GCG	CAG	GCC

Modified E. coli ompA signal sequence with C-terminal AYA and NdeI restriction site (= CATATG):

M K K T A I A I A V A L A G F A T V A Y A

ATGAAAAAAACCGCCATTGCCATTGCCGTGGCCCTGGCAGGCTTTGCCACCGTGG**CATAT G**CC

Alternatively the DNA sequence includes:

ATGAAAAAAACCGCCATTGCCATTGCCGTGGCCCTGGCAGGCTTTGCCACCGTGG**CATAT**
GCG

(SEQ ID NOS 14, 13 and 15 respectively, in order of appearance)

In order to allow an easy switch from E. coli expressed Fab to mammalian expressed IgG formats, the human leader sequences for the IgG light chain (human kappa leader) and IgG heavy chain (human heavy chain leader) were generated to contain the same restriction sites as the C-termini of the ompA (NdeI (AYA)) and phoA (NheI (VLA)) signal sequences. The wildtype and modified human heavy chain leader and human kappa leader sequences are shown in Table 3.

Table 3

Heavy chain leader

A)

Wildtype human heavy chain leader (C-terminal amino acid sequence from position -3 to -1 is VLS without restriction site):

M K H L W F F L L L V A A P R W V L S

ATGAAACACCTGTGGTTCTTCCTCCTGCTGGTGGCAGCTCCCAGATGGGTCCTGTCC

(SEQ ID NOS 17 and 16, respectively, in order of appearance)

Wild type Heavy chain leader C-terminus	V	L	S
	GTC	CTG	TCC

B)

Modified human heavy chain leader with C-terminal VLA and NheI restriction site (= GCTAGC):

M K H L W F F L L L V A A P R W V L A

ATGAAGCACCTGTGGTTCTTTCTGCTGCTGGTGGCCGCTCCCCGGTGGGT**GCTAGC**C

(SEQ ID NOS 19 and 18, respectively, in order of appearance)

C)

Wildtype human kappa leader (C-terminal amino acid sequence from position -3 to -1 is AYG without restriction site):

M V L Q T Q V F I S L L L W I S G A Y G

ATGGTGTTGCAGACCCAGGTCTTCATTTCTCTGTTGCTCTGGATCTCTGGTGCCTACGGG

(SEQ ID NOS 21 and 20, respectively, in order of appearance)

Kappa leader C-terminus	A	Y	G
	GCC	TAC	GGG

D)

Modified human kappa leader with C-terminal AYA and NdeI restriction site (= CATATG):

M V L Q T Q V F I S L L L W I S G A Y A

ATGGTGCTCCAGACCCAGGTGTTCATCAGCCTGCTGCTGTGGATCAGCGGCG**CATATG**CG

(SEQ ID NOS 23 and 22, respectively, in order of appearance)

The selected modified prokaryotic signal sequences and human leader sequences (a) result in high yields of Fab and IgG protein according to the vector system used, (b) provide full compatibility for switching antibody formats, vectors and expression systems between prokaryotic and mammalian systems and (c) are located in the signal/leader sequences thereby maintaining the full germline sequences of FR1.

Example 2: Identification of the most abundant VH/VL pairs in the human repertoire

In its most general sense, the inventors began with the idea that an antibody collection that imitates the human immune system in essential ways may be advantageous. The inventors worked from their hypothesis that the variable heavy chain and variable light chain germline gene pairs abundantly expressed in the human immune repertoire likely have favorable biophysical properties that would lead to more efficient clinical development and increase the safety and efficacy of the resulting antibodies in patients. In order to prove this hypothesis, the first step was to identify the variable heavy chain and variable light chain germline gene pairs prominently expressed in the human immune repertoire.

Example 2.1: Determination of VH/VL pair germline gene usage

In order to identify the predominantly expressed VH/VL germline gene pairs from the human immune repertoire, publically available data was analyzed and human B cells were sampled. As a first step, publically available data was reviewed to identify articles describing the VH/VL germline gene pairs isolated from human B cells. As mentioned, many publically available databases provide antibody sequences, however, many provide only the sequences of either variable domain, VH or VL, but seldom provide the linkage of VH/VL germline gene pairs. The following articles were identified and analyzed in detail: Wardemann H. et al. (2003) Science 301, 1374-1377 and any supporting tables; Yurasov S. et al. (2005) J. Exp. Med. 201, 703-712 and any supporting tables; Tsuiji M. et al. (2006) J. Exp. Med. 203, 393-401 and any supporting tables; Yurasov S. et al. (2006) J. Exp. Med. 203, 2255-2262 and any supporting tables, Tiller T. et al. (2007) Immunity 26, 205-213 and any supporting tables, and Mietzner B. et al. (2008) PNAS 105, 9727-9732 and any supporting tables, all of which are incorporated by

reference in their entireties. Additional VH/VL pair data was identified from a sample of human B cells, as described below.

Example 2.2: Determination of VH/VL pair gene usage from a human sample

In order to obtain additional VH/VL germline gene pair usage data, PBMCs were isolated from a human host. The PBMCs were sorted, the cDNAs of the B cells were amplified using PCR, the DNA from the B cells was sequenced and then the sequences were blasted with IgBLAST (NCBI) to identify the VH/VL germline gene pairs from each B cell.

General methods of isolating and sorting human PBMCs from venous blood and mononuclear cells from bone marrow are described in Tiller et al., J Immunol Methods, 2008 Jan 1;329(1-2):112-24, which is incorporated by reference in its entirety. The PBMCs were isolated and then single sorted according to the cell surface marker of the phenotype of interest. Ig gene transcripts of the single sorted mature naïve (mn) B cells and antibody secreting cells (asc) were then PCR amplified for determination of the VH/VL germline gene pairings. General methods of PCR amplifying cDNA of B cells and the primers useful for the same are also described in Tiller et.al. 2008 (citation above). The specific primers used are shown in Table 4.

Table 4 (SEQ ID NOS 24-60, respectively, in order of appearance):

for μ or γ heavy chain PCR:

HC 1st PCR		
5' L-VH 1	ACAGGTGCCCACTCCCAGGTGCAG	24
5' L-VH 3	AAGGTGTCCAGTGTGARGTGCAG	23
5' L-VH 4/6	CCCAGATGGGTCCTGTCCCAGGTGCAG	27
5' L-VH 5	CAAGGAGTCTGTTCCGAGGTGCAG	24
3' Cμ CH1 (mu)	GGGAATTCTCACAGGAGACGA	21
3' Cg CH1 (gamma)	GGAAGGTGTGCACGCCGCTGGTC	23
HC 2nd PCR		
5' AgeI VH1	CTGCAACCGGTGTACATTCCCAGGTGCAGCTGGTGCAG	38
5' AgeI VH1/5	CTGCAACCGGTGTACATTCCGAGGTGCAGCTGGTGCAG	38
5' AgeI VH3	CTGCAACCGGTGTACATTCTGAGGTGCAGCTGGTGGAG	38
5' AgeI VH3-23	CTGCAACCGGTGTACATTCTGAGGTGCAGCTGTTGGAG	38
5' AgeI VH4	CTGCAACCGGTGTACATTCCCAGGTGCAGCTGCAGGAG	38
5' AgeI VH 4-34	CTGCAACCGGTGTACATTCCCAGGTGCAGCTACAGCAGTG	40
3' SalI JH 1/2/4/5	TGCGAAGTCGACGCTGAGGAGACGGTGACCAG	32
3' SalI JH 3	TGCGAAGTCGACGCTGAAGAGACGGTGACCATTG	34
3' SalI JH 6	TGCGAAGTCGACGCTGAGGAGACGGTGACCGTG	33
3' IgG (internal)	GTTCGGGGAAGTAGTCCTTGAC	22

for kappa light chain PCR:

k LC 1st PCR		
5' L-Vk 1/2	ATGAGGSTCCCYGCTCAGCTGCTGG	25
5' L-Vk 3	CTCTTCCTCCTGCTACTCTGGCTCCCAG	28
5' L-Vk 4	ATTTCTCTGTTGCTCTGGATCTCTG	25
3' Ck 543	GTTTCTCGTAGTCTGCTTTGCTCA	24
k LC 2nd PCR		
5' Pan Vk	ATGACCCAGWCTCCABYCWCCCTG	24
3' Ck 494	GTGCTGTCCTTGCTGTCCTGCT	22

for lambda light chain PCR:

(LC 1st PCR		
5' L-Vl 1	GGTCCTGGGCCCAGTCTGTGCTG	23
5' L-Vl 2	GGTCCTGGGCCCAGTCTGCCCTG	23
5' L-Vl 3	GCTCTGTGACCTCCTATGAGCTG	23
5' L-Vl 4/5	GGTCTCTCTCSCAGCYTGTGCTG	23
5' L-Vl 6	GTTCTTGGGCCAATTTTATGCTG	23
5' L-Vl 7	GGTCCAATTCYCAGGCTGTGGTG	23
5' L-Vl 8	GAGTGGATTCTCAGACTGTGGTG	23
3' C(	CACCAGTGTGGCCTTGTTGGCTTG	24
(LC 2nd PCR		
5'AgeI Vl 1	CTGCTACCGGTTCCTGGGCCCAGTCTGTGCTGACKCAG	38
5'AgeI Vl 2	CTGCTACCGGTTCCTGGGCCCAGTCTGCCCTGACTCAG	38
5'AgeI Vl 3	CTGCTACCGGTTCTGTGACCTCCTATGAGCTGACWCAG	38
5'AgeI Vl 4/5	CTGCTACCGGTTCTCTCTCSCAGCYTGTGCTGACTCA	37
5'AgeI Vl 6	CTGCTACCGGTTCTTGGGCCAATTTTATGCTGACTCAG	38
5'AgeI Vl 7/8	CTGCTACCGGTTCCAATTCYCAGRCTGTGGTGACYCAG	38
3' XhoI Cl	CTCCTCACTCGAGGGYGGGAACAGAGTG	28

cDNAs of the single sorted mature naïve (mn) B cells and antibody secreting cells (asc) were synthesized. Nested PCR was conducted, where human IgH, Igk and IgL V gene transcripts were PCR amplified independently. The sequencing results were blasted with IgBLAST (NCBI) to identify the respective VH, VK, and VL germline genes.

Example 2.3 VH/VL Germline Gene Pairs identified in the human immune repertoire

The VH/VL germline gene pair data identified from the publically available literature as described in Example 2.1 was pooled with the data identified from a human sample as described in Example 2.2. The pooled data was analyzed and is shown as a ranking in Table 6, i.e. the ranking of the percentage/proportion (%) of the VH/VL germline gene pairs identified in the human immune repertoire.

Example 3: Determining the VH and VL germline gene usage

A review of Table 6 shows that a small number of VH/VL pairs are dominant in the human immune repertoire as compared to the total number of germline genes. Wildt et al. at

895-896 described this phenomenon. Wildt et al. also described that the frequently expressed heavy and light chain gene segments are often paired, and observed that half of the pairings sampled corresponded to only five VH/VL germline gene pairs.

Additionally, the pooled data and additional references were evaluated to identify the VH, Vκ, and Vλ germline genes that are independently expressed (not as pairs) in the human immune repertoire. The additional literature references, which include unpaired VH and/or VL germline gene expression, were Brezinschek H.P. et al. (1997) J. Clin. Invest. 99, 2488, Demaison C. et al. (1995) Immunogenetics 42, 342, and Foster S.J. et al. (1997) J. Clin. Invest. 99, 1614, which are both incorporated by reference in their entireties. The data from Examples 2.1 and 2.2 and additional references were pooled and ranked to determine the VH, Vκ, and Vλ germline genes most prominently expressed in the human immune repertoire. The ranking is shown inTable 5.

In comparing Table 5, showing the unlinked VH, Vλ and Vκ germline gene prevalence in the human immune repertoire and Table 6, showing the linked VH/VL pair germline gene prevalence within the human immune repertoire, it was apparent that many of the VH, Vλ and Vκ germline genes that are highly represented when evaluated independent of linkage or pairing were also highly represented in the VH/VL pairings.

This observation is confirmed by the plots shown in Figs. 4-5, which show the VH/VL germline gene pairs of the human immune repertoire. The figures show the actual number of each VH/VL germline gene pair identified from the pooled data, plotted on a matrix, where the Y axis includes the ranking of the VH germline genes, and the X axis includes the ranking of the VL germline genes.

Example 4: Selecting the VH/VL germline gene pairings for further evaluation of their biophysical properties

As a next step, it had to be determined which germline protein pairs were to be tested, as there are ~2500 pairs in the human immune repertoire and the inventors goal was to identify which of the germline protein pairs comprise favorable biophysical properties which would aid in selection and development. One way would be to to test the variable heavy chain and variable light chain germline protein pairs that occur most prominently in the human immune repertoire, for example see Table 6. One could, for example, select the top four hundred pairs for testing, or select the variable heavy chain and variable light chain germline protein pairs present above a certain threshold number. This approach would require the synthesis and testing of a large

number of different variable heavy chain and variable light chain germline protein pair sequences; therefore, such an approach may not be very efficient.

As an alternative approach, the inventors selected a subset of the variable heavy chain and variable light chain germline pairs that are representative of, accurately reproduce, or cover the majority of the prominent pairs from the human immune repertoire. This approach was based, in part, upon the above observation that a small number of variable heavy, variable κ light chain, and variable λ light chain germline genes (unpaired) are dominant in the human immune repertoire. Therefore, a small number of the prominent heavy and light chain germline genes (unpaired) can be combined to generate a group of VH/VL pairs that are representative of the human immune repertoire.

This approach was undertaken in the following way. In Example 3, the variable heavy chain, variable κ light chain, and variable λ light chain germline gene expression was determined. As a next step, an *in silico* analysis was completed of the prominent VH, Vλ and Vκ germline genes, where at least the following factors were evaluated: CDR length, isoelectric point (pI) (the preferred isoelectric point is 7.5 or above as this is should provide stability in a standard pH 5.5 to pH 7 formulation buffer), potential post translational modification sites (PTM's) (specifically, N-linked glycosylation sites (NxS or NxT) or chemical modifications such as Asp cleavage (often at a DP or DQ), Asp isomerization (DS, DG), deamidation (NS, NG) which can occur *in vivo* (in serum) or upon storage in formulation buffer and lead to loss of antibody binding), the presence of Methionines in the CDRs (can be oxidized when exposed to solvent), the presence of unpaired cysteines (will form disulfide bonds with any other unpaired cysteine, thus leading to crosslinking of proteins and/or lower expression yield), deviations from germline, the presence of possible T-cell epitopes, and theoretical aggregation propensity. Selected data from the *in silico* analysis is shown in Figs. 2-3.

Based upon the *in silico* analysis of the most prominent VH, Vλ and Vκ germline genes, a subset of these were selected for synthesis, combination and subsequent functional testing. This subset is shown in Figs. 2-3. When comparing Table 5 and Figs. 2-3,it is clear that not all of the most prominent VH, Vλ and Vκ germline genes were selected for further testing. Of the most prominent VH germline genes, shown in Table 5, IGHV4-34, IGHV4-59, and IGHV3-9 were not selected. Instead, see in Figs. 2-3, IGHV3-74, IGHV3-73, and IGHV6-1 were selected. In total, 20 VH germline genes were selected. Of the most prominent Vκ germline genes, shown in Table 5, IGKV4-1, IGKV2-28/2D-28, IGKV1-33/1D-33, and IGKV1-8 were not selected. In total, 12 Vκ germline genes were selected. Of the most prominent Vλ germline genes shown in Table 5, IGLV1-44 was not selected. In total, 8 Vλ germline genes were selected.

Table 5 shows the ranking of the VH, Vκ, and Vλ germline gene usage from the human immune repertoire and **bolds and underlines** the germline genes that were selected for further functional testing.

Table 5

VH

n=2463

1	**IGHV3-23**	**10,6**
2	**IGHV3-30**	**8,0**
3	**IGHV4-39**	**7,6**
4	IGHV4-34	6,8
5	IGHV4-59	5,8
6	**IGHV1-69**	**5,3**
7	**IGHV5-51**	**4,6**
8	**IGHV3-7**	**4,5**
9	**IGHV1-18**	**4,1**
10	**IGHV3-48**	**4,0**
11	**IGHV3-15**	**3,3**
12	**IGHV3-21**	**3,3**
13	**IGHV1-2**	**3,2**
14	**IGHV3-33**	**3,0**
15	**IGHV4-31**	**3,0**
16	**IGHV3-53**	**2,7**
17	**IGHV3-11**	**2,6**
18	IGHV3-9	2,2
19	**IGHV4-4**	**2,1**
20	**IGHV1-46**	**2,1**
21	**IGHV3-74**	**1,6**
22	IGHV1-24	1,1
23	IGHV4-61	1,1
24	IGHV1-8	1,1
25	IGHV1-3	1,0
26	IGHV3-49	1,0
27	IGHV3-43	0,6
28	IGHV4-28	0,6
29	IGHV3-64	0,5
30	IGHV7-81	0,5
31	IGHV3-13	0,4
32	IGHV3-72	0,4
33	IGHV1-58	0,3
34	**IGHV3-73**	**0,3**
35	IGHV3-66	0,2
36	IGHV7-4.1	0,2
37	IGHV2-5	0,1
38	IGHV4-30.2	0,1

Vκ

n=1656

1	**IGKV3-20**	**16,2**
2	**IGKV1-39/1D-39**	**14,2**
3	**IGKV1-5**	**11,2**
4	**IGKV3-15**	**11,1**
5	IGKV4-1	8,5
6	**IGKV3-11**	**7,6**
7	IGKV2-28/2D-28	6,0
8	IGKV1-33/1D-33	4,6
9	**IGKV2-30**	**2,6**
10	**IGKV1-9**	**2,4**
11	**IGKV1-17**	**2,4**
12	**IGKV1-27**	**2,2**
13	IGKV1-8	1,9
14	**IGKV1-16**	**1,3**
15	**IGKV1-6**	**1,1**
16	**IGKV1-12**	**1,1**
17	IGKV2D-29	1,0
18	IGKV1-13	0,7
19	IGKV1D-8	0,5
20	IGKV2-24	0,5
21	IGKV5-2	0,4
22	IGKV1D-12	0,3
23	IGKV2-40/2D-40	0,3
24	IGKV3D-20	0,3
25	IGKV1D-43	0,2
26	IGKV2D-30	0,2
27	IGKV3D-11	0,2
28	IGKV3D-15	0,2
29	IGKV2-29	0,2
30	IGKV1D-16	0,1
31	IGKV1D-17	0,1
32	IGKV3D-7	0,1
33	IGKV6-21/6D-21	0,1
34	IGKV6D-41	0,1
35	IGKV1D-13	0,0

Vλ

n=780

1	**IGLV2-14**	**18,1**
2	**IGLV1-40**	**11,3**
3	IGLV1-44	11,3
4	**IGLV1-51**	**10,0**
5	**IGLV2-23**	**8,1**
6	**IGLV3-21**	**8,1**
7	**IGLV1-47**	**6,5**
8	**IGLV3-1**	**5,3**
9	**IGLV2-11**	**5,1**
10	IGLV2-8	4,5
11	IGLV6-57	1,7
12	IGLV3-25	1,5
13	IGLV7-46	1,5
14	IGLV1-36	1,2
15	IGLV7-43	1,2
16	IGLV9-49	1,2
17	IGLV4-69	1,0
18	IGLV2-18	0,6
19	IGLV3-10	0,5
20	IGLV3-27	0,5
21	IGLV3-9	0,3
22	IGLV3-12	0,1
23	IGLV3-19	0,1
24	IGLV3-22	0,1
25	IGLV4-60	0,1
26	IGLV8-61	0,1
27	IGLV3-16	0,0
28	IGLV4-3	0,0
29	IGLV5-37	0,0
30	IGLV5-39	0,0
31	IGLV5-45	0,0
32	IGLV5-52	0,0
33	IGLV10-54	0,0

39	IGHV3-20	0,1
40	**IGHV6-1**	**0,0**
41	IGHV1-e	0,0
42	IGHV1-f	0,0
43	IGHV1-45	0,0
44	IGHV2-26	0,0
45	IGHV2-70	0,0
46	IGHV3-d	0,0
47	IGHV4-b	0,0
48	IGHV4-30.4	0,0
49	IGHV5-a	0,0

Example 4.1: Recombination of abundant VH, Vκ, and Vλ germline genes to yield representation of VH/VL most prominent pairs in the human immune repertoire

As a next step, the 20 VH, 12 Vκ and 8 Vλ selected VH, Vκ, and Vλ germline genes were synthesized and combined to generate 400 VH/VL germline gene pairs, which pairs were subsequently tested for their biophysical properties. Table 6 shows that the 400 VH/VL germline gene pairs generated for functional testing do, in fact, accurately reproduce or cover the majority of the prominent VH/VL germline gene pairs in the human immune repertoire. Table 6 shows the ranking of the VH/VL pairs expressed in the human immune repertoire, wherein the 400 VH/VL pairs that were tested are bolded and underlined.

Table 6: The 400 VH/VL germline gene pairs functionally tested are representative of the VH/VL germline gene pairs identified in the human immune repertoire

"pos": represents the position of relative ranking of the VH/VL pairs as determined by the percentage (%) of each VH/VL pair from the total pooled data.

N= 2137 B cells

pos	**V heavy**	**V light**	**%**
1	**IGHV3-23**	**IGKV1-5**	**1,26**
2	IGHV4-34	IGKV3-20	1,17
3	**IGHV3-23**	**IGKV3-20**	**1,12**
4	**IGHV4-39**	**IGKV3-15**	**1,03**
5	**IGHV3-23**	**IGKV3-15**	**0,94**
6	IGHV4-59	IGKV1-39/1D-39	0,89
7	**IGHV4-39**	**IGKV1-39/1D-39**	**0,84**
	IGHV4-34	IGKV1-39/1D-39	0,84
8	IGHV4-59	IGKV3-20	0,70
	IGHV1-18	**IGKV3-20**	**0,70**
9	**IGHV3-30**	**IGKV3-20**	**0,66**
	IGHV4-39	**IGKV1-5**	**0,66**
	IGHV1-69	**IGKV1-39/1D-39**	**0,66**

	IGHV5-51	**IGLV 1-40**	**0,66**
10	IGHV3-23	IGKV4-1	0,61
	IGHV4-39	**IGKV3-20**	**0,61**
	IGHV3-23	**IGLV 2-14**	**0,61**
	IGHV4-39	**IGLV 3-21**	**0,61**
11	**IGHV3-23**	**IGKV1-39/1D-39**	**0,56**
	IGHV3-30	**IGKV1-39/1D-39**	**0,56**
	IGHV3-30	**IGKV3-11**	**0,56**
	IGHV1-69	**IGKV3-20**	**0,56**
	IGHV3-48	**IGKV3-20**	**0,56**
	IGHV1-2	**IGKV3-20**	**0,56**
12	IGHV3-30	IGKV4-1	0,51
	IGHV5-51	**IGLV 2-14**	**0,51**
13	IGHV4-59	IGKV4-1	0,47
	IGHV5-51	**IGKV3-20**	**0,47**
	IGHV3-7	**IGKV1-39/1D-39**	**0,47**
	IGHV3-7	**IGKV1-5**	**0,47**
	IGHV3-15	**IGKV3-20**	**0,47**
	IGHV4-39	**IGLV 2-14**	**0,47**
	IGHV4-39	IGLV 2-8	0,47
	IGHV4-34	IGLV 2-14	0,47
14	**IGHV3-23**	**IGKV3-11**	**0,42**
	IGHV3-30	**IGKV1-5**	**0,42**
	IGHV3-30	**IGKV3-15**	**0,42**
	IGHV4-34	IGKV1-5	0,42
	IGHV3-21	**IGKV1-5**	**0,42**
	IGHV3-21	**IGKV3-15**	**0,42**
	IGHV3-30	**IGLV 1-51**	**0,42**
	IGHV4-34	IGLV 1-51	0,42
	IGHV3-21	**IGLV 1-51**	**0,42**
	IGHV3-53	IGLV 1-44	0,42
15	IGHV4-59	IGKV3-15	0,37
	IGHV4-34	IGKV3-15	0,37
	IGHV5-51	IGKV4-1	0,37
	IGHV1-69	IGKV4-1	0,37
	IGHV1-69	**IGKV3-11**	**0,37**
	IGHV3-7	**IGKV3-15**	**0,37**
	IGHV1-18	**IGKV1-39/1D-39**	**0,37**
	IGHV3-48	**IGKV1-39/1D-39**	**0,37**
	IGHV3-33	**IGKV3-15**	**0,37**
	IGHV3-53	**IGKV1-5**	**0,37**
	IGHV4-59	IGLV 1-40	0,37
	IGHV1-69	**IGLV 2-14**	**0,37**
	IGHV1-69	IGLV 1-44	0,37
	IGHV4-31	**IGLV 2-14**	**0,37**
	IGHV1-2	**IGLV 2-14**	**0,37**
16	IGHV3-23	IGKV2-28/2D-28	0,33
	IGHV3-30	**IGKV1-9**	**0,33**

	IGHV4-34	IGKV4-1	0,33
	IGHV5-51	**IGKV1-39/1D-39**	**0,33**
	IGHV5-51	**IGKV3-15**	**0,33**
	IGHV1-69	**IGKV3-15**	**0,33**
	IGHV1-18	IGKV1-33/1D-33	0,33
	IGHV3-48	**IGKV3-11**	**0,33**
	IGHV3-21	**IGKV1-39/1D-39**	**0,33**
	IGHV4-31	**IGKV3-20**	**0,33**
	IGHV4-31	**IGKV3-11**	**0,33**
	IGHV3-30	**IGLV 2-14**	**0,33**
	IGHV4-39	IGLV 1-44	0,33
	IGHV1-69	**IGLV 1-40**	**0,33**
	IGHV3-9	IGLV 2-23	0,33
17	IGHV3-23	IGKV1-33/1D-33	0,28
	IGHV4-39	**IGKV3-11**	**0,28**
	IGHV4-34	IGKV3-11	0,28
	IGHV4-34	IGKV2-28/2D-28	0,28
	IGHV5-51	**IGKV3-11**	**0,28**
	IGHV5-51	IGKV1-13	0,28
	IGHV3-7	**IGKV3-20**	**0,28**
	IGHV3-48	**IGKV3-15**	**0,28**
	IGHV3-48	IGKV4-1	0,28
	IGHV3-48	IGKV1-33/1D-33	0,28
	IGHV3-15	**IGKV1-39/1D-39**	**0,28**
	IGHV3-15	**IGKV1-5**	**0,28**
	IGHV1-2	**IGKV1-39/1D-39**	**0,28**
	IGHV3-33	**IGKV3-20**	**0,28**
	IGHV3-33	**IGKV1-39/1D-39**	**0,28**
	IGHV3-33	IGKV4-1	0,28
	IGHV3-53	**IGKV3-15**	**0,28**
	IGHV3-11	**IGKV1-5**	**0,28**
	IGHV4-4	**IGKV3-20**	**0,28**
	IGHV1-46	**IGKV3-20**	**0,28**
	IGHV3-23	**IGLV 1-40**	**0,28**
	IGHV3-23	**IGLV 3-21**	**0,28**
	IGHV4-39	**IGLV 1-40**	**0,28**
	IGHV4-34	IGLV 1-40	0,28
	IGHV4-34	IGLV 1-47	0,28
	IGHV3-48	**IGLV 2-14**	**0,28**
	IGHV3-48	**IGLV 1-47**	**0,28**
	IGHV1-2	**IGLV 1-40**	**0,28**
	IGHV3-9	IGLV 2-14	0,28
	IGHV4-4	IGLV 1-44	0,28
18	**IGHV3-23**	**IGKV1-17**	**0,23**
	IGHV4-39	IGKV4-1	0,23
	IGHV4-39	IGKV2-28/2D-28	0,23
	IGHV1-69	**IGKV1-5**	**0,23**
	IGHV3-7	IGKV4-1	0,23

	IGHV1-18	**IGKV1-5**	**0,23**
	IGHV1-18	IGKV2-28/2D-28	0,23
	IGHV3-21	**IGKV3-20**	**0,23**
	IGHV3-33	**IGKV1-5**	**0,23**
	IGHV3-53	**IGKV1-39/1D-39**	**0,23**
	IGHV3-53	IGKV1-33/1D-33	0,23
	IGHV3-11	**IGKV1-39/1D-39**	**0,23**
	IGHV3-11	**IGKV3-15**	**0,23**
	IGHV4-4	**IGKV1-39/1D-39**	**0,23**
	IGHV1-46	**IGKV1-39/1D-39**	**0,23**
	IGHV4-61	IGKV4-1	0,23
	IGHV3-23	IGLV 1-44	0,23
	IGHV3-23	**IGLV 2-11**	**0,23**
	IGHV3-23	**IGLV 3-1**	**0,23**
	IGHV3-30	**IGLV 1-40**	**0,23**
	IGHV4-39	**IGLV 1-51**	**0,23**
	IGHV4-39	**IGLV 2-23**	**0,23**
	IGHV4-59	IGLV 3-1	0,23
	IGHV5-51	IGLV 1-44	0,23
	IGHV1-69	**IGLV 1-51**	**0,23**
	IGHV1-69	**IGLV 2-11**	**0,23**
	IGHV1-18	**IGLV 2-14**	**0,23**
	IGHV1-18	**IGLV 1-40**	**0,23**
	IGHV3-21	**IGLV 2-14**	**0,23**
	IGHV1-2	IGLV 1-44	0,23
19	**IGHV3-23**	**IGKV1-27**	**0,19**
	IGHV3-23	IGKV1-8	0,19
	IGHV3-30	IGKV2-28/2D-28	0,19
	IGHV4-39	IGKV1-33/1D-33	0,19
	IGHV4-39	**IGKV1-27**	**0,19**
	IGHV4-59	IGKV3-11	0,19
	IGHV5-51	**IGKV1-5**	**0,19**
	IGHV5-51	IGKV2-28/2D-28	0,19
	IGHV3-7	**IGKV3-11**	**0,19**
	IGHV3-7	**IGKV2-30**	**0,19**
	IGHV1-18	**IGKV3-15**	**0,19**
	IGHV1-18	**IGKV3-11**	**0,19**
	IGHV3-21	IGKV4-1	0,19
	IGHV3-15	**IGKV3-15**	**0,19**
	IGHV3-15	IGKV4-1	0,19
	IGHV3-15	IGKV1-33/1D-33	0,19
	IGHV4-31	**IGKV1-39/1D-39**	**0,19**
	IGHV4-31	**IGKV1-5**	**0,19**
	IGHV4-31	**IGKV3-15**	**0,19**
	IGHV4-31	IGKV2-28/2D-28	0,19
	IGHV3-33	IGKV2-28/2D-28	0,19
	IGHV3-53	IGKV4-1	0,19
	IGHV3-53	**IGKV3-11**	**0,19**

	IGHV3-74	**IGKV3-20**	**0,19**
	IGHV4-4	**IGKV1-5**	**0,19**
	IGHV1-46	**IGKV1-9**	**0,19**
	IGHV1-8	IGKV3-15	0,19
	IGHV1-24	IGKV3-11	0,19
	IGHV1-3	IGKV1-39/1D-39	0,19
	IGHV3-49	IGKV1-39/1D-39	0,19
	IGHV3-23	**IGLV 2-23**	**0,19**
	IGHV3-30	IGLV 1-44	0,19
	IGHV4-59	IGLV 2-14	0,19
	IGHV4-59	IGLV 1-44	0,19
	IGHV4-59	IGLV 1-51	0,19
	IGHV4-34	IGLV 2-8	0,19
	IGHV5-51	**IGLV 1-47**	**0,19**
	IGHV1-69	IGLV 2-8	0,19
	IGHV3-7	**IGLV 1-40**	**0,19**
	IGHV3-15	IGLV 1-44	0,19
	IGHV4-31	**IGLV 2-23**	**0,19**
	IGHV3-33	**IGLV 2-14**	**0,19**
	IGHV3-33	**IGLV 1-47**	**0,19**
	IGHV3-33	**IGLV 2-23**	**0,19**
	IGHV3-33	**IGLV 3-21**	**0,19**
	IGHV3-9	IGLV 1-44	0,19
	IGHV4-4	**IGLV 2-14**	**0,19**
	IGHV1-46	**IGLV 1-51**	**0,19**
	IGHV4-61	IGLV 1-44	0,19
	IGHV1-8	IGLV 2-14	0,19
	IGHV4-28	IGLV 2-23	0,19
20	**IGHV3-23**	**IGKV1-9**	**0,14**
	IGHV3-23	**IGKV1-16**	**0,14**
	IGHV4-39	**IGKV1-6**	**0,14**
	IGHV4-59	IGKV1-5	0,14
	IGHV4-59	IGKV1-27	0,14
	IGHV4-34	IGKV1-33/1D-33	0,14
	IGHV5-51	IGKV1-33/1D-33	0,14
	IGHV1-69	IGKV2-28/2D-28	0,14
	IGHV1-69	IGKV1-33/1D-33	0,14
	IGHV3-7	IGKV2-28/2D-28	0,14
	IGHV3-7	IGKV1-8	0,14
	IGHV3-48	IGKV2-28/2D-28	0,14
	IGHV3-48	IGKV1-8	0,14
	IGHV3-15	**IGKV3-11**	**0,14**
	IGHV3-15	IGKV2-28/2D-28	0,14
	IGHV3-15	**IGKV1-9**	**0,14**
	IGHV4-31	IGKV1-33/1D-33	0,14
	IGHV1-2	**IGKV1-5**	**0,14**
	IGHV1-2	IGKV4-1	0,14
	IGHV3-11	**IGKV3-20**	**0,14**

	IGHV3-11	**IGKV3-11**	**0,14**
	IGHV3-11	IGKV2-28/2D-28	0,14
	IGHV3-9	IGKV1-39/1D-39	0,14
	IGHV3-9	IGKV1-5	0,14
	IGHV3-9	IGKV4-1	0,14
	IGHV3-9	IGKV2D-29	0,14
	IGHV3-74	**IGKV1-39/1D-39**	**0,14**
	IGHV3-74	**IGKV1-5**	**0,14**
	IGHV3-74	**IGKV3-15**	**0,14**
	IGHV3-74	IGKV4-1	0,14
	IGHV4-4	**IGKV3-15**	**0,14**
	IGHV4-4	IGKV4-1	0,14
	IGHV4-4	**IGKV3-11**	0,14
	IGHV1-46	**IGKV1-5**	0,14
	IGHV1-46	**IGKV3-15**	0,14
	IGHV4-61	IGKV1-39/1D-39	0,14
	IGHV1-24	IGKV1-39/1D-39	0,14
	IGHV1-24	IGKV3-15	0,14
	IGHV1-3	IGKV3-15	0,14
	IGHV3-49	IGKV1-17	0,14
	IGHV3-43	IGKV1-5	0,14
	IGHV7-81	IGKV3-20	0,14
	IGHV3-13	IGKV1-39/1D-39	0,14
	IGHV3-23	**IGLV 1-51**	0,14
	IGHV3-30	**IGLV 3-21**	0,14
	IGHV3-30	**IGLV 3-1**	0,14
	IGHV4-39	**IGLV 1-47**	0,14
	IGHV4-39	IGLV 2-18	0,14
	IGHV4-59	IGLV 1-47	0,14
	IGHV5-51	**IGLV 2-23**	0,14
	IGHV5-51	**IGLV 3-21**	0,14
	IGHV1-69	**IGLV 2-23**	0,14
	IGHV3-7	IGLV 1-44	0,14
	IGHV3-7	**IGLV 1-51**	0,14
	IGHV3-7	**IGLV 1-47**	0,14
	IGHV3-7	**IGLV 3-21**	0,14
	IGHV1-18	IGLV 1-44	0,14
	IGHV1-18	IGLV 1-51	0,14
	IGHV3-48	**IGLV 3-1**	0,14
	IGHV3-21	**IGLV 1-47**	0,14
	IGHV3-15	IGLV 7-46	0,14
	IGHV4-31	**IGLV 1-40**	0,14
	IGHV4-31	**IGLV 1-51**	0,14
	IGHV4-31	**IGLV 1-47**	0,14
	IGHV1-2	**IGLV 1-51**	0,14
	IGHV1-2	**IGLV 2-23**	0,14
	IGHV1-2	**IGLV 3-1**	0,14
	IGHV3-11	**IGLV 2-14**	0,14

	IGHV3-11	IGLV 1-44	0,14
	IGHV3-11	**IGLV 2-11**	0,14
	IGHV3-11	**IGLV 3-1**	0,14
	IGHV3-9	IGLV 1-47	0,14
	IGHV3-9	IGLV 2-11	0,14
	IGHV3-74	**IGLV 2-23**	0,14
	IGHV3-74	**IGLV 3-21**	0,14
	IGHV4-4	**IGLV 1-40**	0,14
	IGHV1-46	**IGLV 2-14**	0,14
	IGHV1-46	IGLV 1-44	0,14
	IGHV4-61	IGLV 2-14	0,14
21	IGHV3-23	IGKV2D-29	0,09
	IGHV3-23	IGKV2-29	0,09
	IGHV3-23	IGKV2-40/2D-40	0,09
	IGHV3-30	IGKV1-33/1D-33	0,09
	IGHV3-30	**IGKV2-30**	0,09
	IGHV3-30	IGKV1-8	0,09
	IGHV3-30	**IGKV1-6**	0,09
	IGHV3-30	IGKV2-24	0,09
	IGHV3-30	IGKV1D-8	0,09
	IGHV4-39	**IGKV2-30**	0,09
	IGHV4-59	IGKV1-33/1D-33	0,09
	IGHV4-59	IGKV1-12	0,09
	IGHV4-34	IGKV1-9	0,09
	IGHV4-34	IGKV1-17	0,09
	IGHV4-34	IGKV1-16	0,09
	IGHV5-51	**IGKV2-30**	0,09
	IGHV1-69	**IGKV1-27**	0,09
	IGHV1-69	IGKV1-8	0,09
	IGHV1-69	IGKV3D-15	0,09
	IGHV3-7	**IGKV1-9**	0,09
	IGHV3-7	**IGKV1-17**	0,09
	IGHV3-7	**IGKV1-27**	0,09
	IGHV3-7	IGKV1-13	0,09
	IGHV1-18	IGKV4-1	0,09
	IGHV1-18	**IGKV2-30**	0,09
	IGHV3-48	**IGKV1-9**	0,09
	IGHV3-48	**IGKV1-17**	0,09
	IGHV3-48	**IGKV1-16**	0,09
	IGHV3-21	**IGKV3-11**	0,09
	IGHV3-21	IGKV2-28/2D-28	0,09
	IGHV3-21	**IGKV1-27**	0,09
	IGHV3-21	IGKV1-8	0,09
	IGHV3-21	**IGKV1-6**	0,09
	IGHV4-31	IGKV4-1	0,09
	IGHV4-31	**IGKV1-17**	0,09
	IGHV4-31	**IGKV1-27**	0,09
	IGHV1-2	**IGKV3-15**	0,09

	IGHV1-2	IGKV2-28/2D-28	0,09
	IGHV1-2	**IGKV1-27**	0,09
	IGHV3-33	**IGKV3-11**	0,09
	IGHV3-33	IGKV1-33/1D-33	0,09
	IGHV3-33	**IGKV1-9**	0,09
	IGHV3-53	**IGKV3-20**	0,09
	IGHV3-53	**IGKV1-27**	0,09
	IGHV3-53	IGKV1-8	0,09
	IGHV3-11	IGKV4-1	0,09
	IGHV3-11	**IGKV1-6**	0,09
	IGHV3-9	IGKV3-15	0,09
	IGHV3-9	IGKV3-11	0,09
	IGHV3-9	IGKV1-16	0,09
	IGHV3-74	**IGKV3-11**	0,09
	IGHV3-74	**IGKV2-30**	0,09
	IGHV4-4	IGKV2-28/2D-28	0,09
	IGHV4-4	IGKV2D-29	0,09
	IGHV1-46	**IGKV3-11**	0,09
	IGHV1-46	**IGKV1-27**	0,09
	IGHV1-46	**IGKV1-16**	0,09
	IGHV4-61	IGKV3-15	0,09
	IGHV1-8	IGKV3-20	0,09
	IGHV1-8	IGKV4-1	0,09
	IGHV1-24	IGKV2-28/2D-28	0,09
	IGHV1-24	IGKV2-30	0,09
	IGHV1-3	IGKV3-20	0,09
	IGHV3-49	IGKV3-20	0,09
	IGHV3-49	IGKV1-5	0,09
	IGHV3-43	IGKV3-11	0,09
	IGHV3-64	IGKV1-5	0,09
	IGHV3-64	IGKV3-11	0,09
	IGHV7-81	IGKV1-39/1D-39	0,09
	IGHV3-13	IGKV4-1	0,09
	IGHV3-72	IGKV1-5	0,09
	IGHV3-72	IGKV3-15	0,09
	IGHV1-58	IGKV3-20	0,09
	IGHV3-66	IGKV1-39/1D-39	0,09
	IGHV3-23	IGLV 1-36	0,09
	IGHV3-30	**IGLV 2-23**	0,09
	IGHV3-30	**IGLV 2-11**	0,09
	IGHV3-30	IGLV 9-49	0,09
	IGHV3-30	IGLV 3-10	0,09
	IGHV4-39	**IGLV 3-1**	0,09
	IGHV4-39	IGLV 6-57	0,09
	IGHV4-59	IGLV 2-23	0,09
	IGHV4-59	IGLV 3-21	0,09
	IGHV4-59	IGLV 2-11	0,09
	IGHV4-34	IGLV 1-44	0,09

	IGHV4-34	IGLV 2-23	0,09
	IGHV4-34	IGLV 3-21	0,09
	IGHV4-34	IGLV 3-25	0,09
	IGHV5-51	IGLV 1-36	0,09
	IGHV5-51	IGLV 3-25	0,09
	IGHV1-69	**IGLV 1-47**	0,09
	IGHV1-69	**IGLV 3-21**	0,09
	IGHV1-69	**IGLV 3-1**	0,09
	IGHV3-7	**IGLV 2-14**	0,09
	IGHV1-18	IGLV 2-8	0,09
	IGHV1-18	IGLV 6-57	0,09
	IGHV3-48	**IGLV 2-11**	0,09
	IGHV3-21	**IGLV 1-40**	0,09
	IGHV3-21	IGLV 1-44	0,09
	IGHV3-21	**IGLV 3-21**	0,09
	IGHV3-21	**IGLV 2-11**	0,09
	IGHV3-21	IGLV 4-69	0,09
	IGHV3-15	**IGLV 1-40**	0,09
	IGHV3-15	**IGLV 1-51**	0,09
	IGHV3-15	**IGLV 3-1**	0,09
	IGHV3-15	IGLV 2-8	0,09
	IGHV3-15	IGLV 7-43	0,09
	IGHV4-31	**IGLV 3-21**	0,09
	IGHV1-2	IGLV 2-8	0,09
	IGHV1-2	IGLV 7-46	0,09
	IGHV3-33	IGLV 6-57	0,09
	IGHV3-53	**IGLV 2-14**	0,09
	IGHV3-11	**IGLV 2-23**	0,09
	IGHV3-11	**IGLV 3-21**	0,09
	IGHV3-11	IGLV 4-69	0,09
	IGHV3-9	IGLV 3-21	0,09
	IGHV3-9	IGLV 2-8	0,09
	IGHV3-74	**IGLV 2-14**	0,09
	IGHV4-4	**IGLV 1-51**	0,09
	IGHV4-4	**IGLV 2-23**	0,09
	IGHV4-4	IGLV 2-8	0,09
	IGHV1-46	**IGLV 2-11**	0,09
	IGHV4-61	IGLV 2-11	0,09
	IGHV1-8	IGLV 1-47	0,09
	IGHV1-24	IGLV 2-23	0,09
	IGHV1-3	IGLV 2-14	0,09
	IGHV1-3	IGLV 2-23	0,09
	IGHV1-3	IGLV 3-1	0,09
	IGHV3-49	IGLV 3-21	0,09
	IGHV4-28	IGLV 1-44	0,09
	IGHV4-28	IGLV 1-51	0,09
	IGHV4-28	IGLV 1-36	0,09
	IGHV3-43	IGLV 1-51	0,09

	IGHV3-64	IGLV 3-21	0,09
	IGHV7-81	IGLV 2-14	0,09
	IGHV7-81	IGLV 3-21	0,09
22	**IGHV3-23**	**IGKV2-30**	0,05
	IGHV3-23	**IGKV1-12**	0,05
	IGHV3-23	IGKV3D-20	0,05
	IGHV3-23	IGKV1D-12	0,05
	IGHV3-23	IGKV1D-13	0,05
	IGHV3-30	**IGKV1-17**	0,05
	IGHV3-30	**IGKV1-27**	0,05
	IGHV3-30	**IGKV1-16**	0,05
	IGHV3-30	IGKV2D-29	0,05
	IGHV3-30	IGKV1-13	0,05
	IGHV3-30	IGKV5-2	0,05
	IGHV3-30	IGKV2D-30	0,05
	IGHV4-39	**IGKV1-17**	0,05
	IGHV4-39	IGKV3D-15	0,05
	IGHV4-59	IGKV2-30	0,05
	IGHV4-59	IGKV1-17	0,05
	IGHV4-59	IGKV1-8	0,05
	IGHV4-59	IGKV1-16	0,05
	IGHV4-59	IGKV1D-43	0,05
	IGHV4-59	IGKV2D-30	0,05
	IGHV4-59	IGKV1D-17	0,05
	IGHV4-34	IGKV1-27	0,05
	IGHV4-34	IGKV1-8	0,05
	IGHV4-34	IGKV1-12	0,05
	IGHV5-51	**IGKV1-9**	0,05
	IGHV5-51	**IGKV1-17**	0,05
	IGHV5-51	**IGKV1-27**	0,05
	IGHV5-51	**IGKV1-12**	0,05
	IGHV1-69	**IGKV2-30**	0,05
	IGHV1-69	**IGKV1-16**	0,05
	IGHV1-69	**IGKV1-6**	0,05
	IGHV1-69	IGKV2D-29	0,05
	IGHV1-69	IGKV2D-30	0,05
	IGHV1-69	IGKV1D-16	0,05
	IGHV3-7	**IGKV1-6**	0,05
	IGHV3-7	IGKV1D-8	0,05
	IGHV3-7	IGKV1D-17	0,05
	IGHV1-18	**IGKV1-17**	0,05
	IGHV1-18	IGKV1-8	0,05
	IGHV1-18	**IGKV1-16**	0,05
	IGHV1-18	**IGKV1-12**	0,05
	IGHV1-18	IGKV1-13	0,05
	IGHV1-18	IGKV2-40/2D-40	0,05
	IGHV3-48	**IGKV1-5**	0,05
	IGHV3-48	**IGKV1-27**	0,05

	IGHV3-48	**IGKV1-6**	0,05
	IGHV3-48	IGKV2D-29	0,05
	IGHV3-48	IGKV3D-20	0,05
	IGHV3-48	IGKV1D-12	0,05
	IGHV3-21	IGKV2D-29	0,05
	IGHV3-15	**IGKV2-30**	0,05
	IGHV3-15	**IGKV1-27**	0,05
	IGHV3-15	IGKV2D-29	0,05
	IGHV3-15	IGKV1-13	0,05
	IGHV3-15	IGKV1D-43	0,05
	IGHV4-31	**IGKV1-6**	0,05
	IGHV4-31	IGKV2-29	0,05
	IGHV4-31	IGKV2-40/2D-40	0,05
	IGHV1-2	IGKV1-33/1D-33	0,05
	IGHV1-2	**IGKV2-30**	0,05
	IGHV1-2	IGKV1-8	0,05
	IGHV1-2	**IGKV1-6**	0,05
	IGHV3-33	**IGKV1-17**	0,05
	IGHV3-33	IGKV1-8	0,05
	IGHV3-33	**IGKV1-16**	0,05
	IGHV3-33	IGKV2-24	0,05
	IGHV3-53	IGKV2-28/2D-28	0,05
	IGHV3-53	**IGKV1-9**	0,05
	IGHV3-53	**IGKV1-17**	0,05
	IGHV3-53	**IGKV1-12**	0,05
	IGHV3-53	IGKV2-29	0,05
	IGHV3-53	IGKV1D-16	0,05
	IGHV3-11	IGKV1-33/1D-33	0,05
	IGHV3-11	**IGKV1-9**	0,05
	IGHV3-11	**IGKV1-17**	0,05
	IGHV3-11	**IGKV1-12**	0,05
	IGHV3-11	IGKV1D-8	0,05
	IGHV3-9	IGKV3-20	0,05
	IGHV3-9	IGKV2-28/2D-28	0,05
	IGHV3-9	IGKV1-17	0,05
	IGHV3-9	IGKV1-27	0,05
	IGHV3-9	IGKV1-8	0,05
	IGHV3-9	IGKV1-12	0,05
	IGHV3-9	IGKV1D-8	0,05
	IGHV4-4	**IGKV1-17**	0,05
	IGHV4-4	**IGKV1-27**	0,05
	IGHV4-4	**IGKV1-6**	0,05
	IGHV4-4	IGKV1D-8	0,05
	IGHV1-46	IGKV4-1	0,05
	IGHV1-46	IGKV1-33/1D-33	0,05
	IGHV1-46	IGKV1-8	0,05
	IGHV4-61	IGKV3-11	0,05
	IGHV4-61	IGKV2-28/2D-28	0,05

	IGHV4-61	IGKV1-16	0,05
	IGHV4-61	IGKV1-12	0,05
	IGHV4-61	IGKV1-13	0,05
	IGHV1-8	IGKV1-39/1D-39	0,05
	IGHV1-8	IGKV1-5	0,05
	IGHV1-8	IGKV3-11	0,05
	IGHV1-8	IGKV2-28/2D-28	0,05
	IGHV1-8	IGKV1-33/1D-33	0,05
	IGHV1-8	IGKV1-9	0,05
	IGHV1-8	IGKV2-29	0,05
	IGHV1-24	IGKV3-20	0,05
	IGHV1-24	IGKV4-1	0,05
	IGHV1-24	IGKV1-33/1D-33	0,05
	IGHV1-24	IGKV2-24	0,05
	IGHV1-24	IGKV2-40/2D-40	0,05
	IGHV1-3	IGKV1-5	0,05
	IGHV1-3	IGKV1-33/1D-33	0,05
	IGHV1-3	IGKV2-30	0,05
	IGHV1-3	IGKV1-6	0,05
	IGHV1-3	IGKV2D-29	0,05
	IGHV3-49	IGKV3-15	0,05
	IGHV3-49	IGKV3-11	0,05
	IGHV3-49	IGKV2-28/2D-28	0,05
	IGHV4-28	IGKV3-20	0,05
	IGHV4-28	IGKV1-39/1D-39	0,05
	IGHV3-43	IGKV3-15	0,05
	IGHV3-43	IGKV4-1	0,05
	IGHV3-43	IGKV2-28/2D-28	0,05
	IGHV3-43	IGKV1-33/1D-33	0,05
	IGHV3-64	IGKV3-15	0,05
	IGHV3-64	IGKV1-9	0,05
	IGHV3-64	IGKV2D-29	0,05
	IGHV7-81	IGKV1-5	0,05
	IGHV7-81	IGKV4-1	0,05
	IGHV7-81	IGKV2-28/2D-28	0,05
	IGHV3-13	IGKV1-5	0,05
	IGHV3-13	IGKV1-33/1D-33	0,05
	IGHV3-13	IGKV1-9	0,05
	IGHV3-13	IGKV2-30	0,05
	IGHV3-72	IGKV3-20	0,05
	IGHV3-72	IGKV1-9	0,05
	IGHV3-72	IGKV1-17	0,05
	IGHV3-72	IGKV1-16	0,05
	IGHV3-73	IGKV2-28/2D-28	0,05
	IGHV3-73	**IGKV1-9**	0,05
	IGHV1-58	IGKV1-5	0,05
	IGHV1-58	IGKV4-1	0,05
	IGHV1-58	IGKV3-11	0,05

	IGHV4-30.2	IGKV1-39/1D-39	0,05
	IGHV4-30.2	IGKV4-1	0,05
	IGHV7-4.1	IGKV1-39/1D-39	0,05
	IGHV7-4.1	IGKV1-5	0,05
	IGHV3-20	IGKV1-39/1D-39	0,05
	IGHV3-23	**IGLV 1-47**	0,05
	IGHV3-23	IGLV 2-8	0,05
	IGHV3-23	IGLV 7-43	0,05
	IGHV3-23	IGLV 2-18	0,05
	IGHV3-23	IGLV 3-19	0,05
	IGHV3-30	**IGLV 1-47**	0,05
	IGHV3-30	IGLV 2-8	0,05
	IGHV3-30	IGLV 6-57	0,05
	IGHV3-30	IGLV 3-27	0,05
	IGHV4-39	IGLV 7-46	0,05
	IGHV4-39	IGLV 3-9	0,05
	IGHV4-59	IGLV 2-8	0,05
	IGHV4-59	IGLV 6-57	0,05
	IGHV4-59	IGLV 3-12	0,05
	IGHV4-34	IGLV 2-11	0,05
	IGHV4-34	IGLV 1-36	0,05
	IGHV4-34	IGLV 7-43	0,05
	IGHV4-34	IGLV 9-49	0,05
	IGHV5-51	IGLV 7-43	0,05
	IGHV1-69	IGLV 6-57	0,05
	IGHV1-69	IGLV 3-25	0,05
	IGHV1-69	IGLV 3-10	0,05
	IGHV3-7	**IGLV 2-23**	0,05
	IGHV3-7	**IGLV 3-1**	0,05
	IGHV3-7	IGLV 2-8	0,05
	IGHV3-7	IGLV 7-46	0,05
	IGHV3-7	IGLV 3-27	0,05
	IGHV1-18	IGLV 2-23	0,05
	IGHV1-18	**IGLV 2-11**	0,05
	IGHV1-18	IGLV 1-36	0,05
	IGHV1-18	IGLV 3-25	0,05
	IGHV1-18	IGLV 3-10	0,05
	IGHV3-48	**IGLV 1-40**	0,05
	IGHV3-48	IGLV 1-44	0,05
	IGHV3-48	**IGLV 1-51**	0,05
	IGHV3-48	**IGLV 2-23**	0,05
	IGHV3-48	**IGLV 3-21**	0,05
	IGHV3-48	IGLV 3-25	0,05
	IGHV3-48	IGLV 7-46	0,05
	IGHV3-48	IGLV 9-49	0,05
	IGHV3-21	**IGLV 2-23**	0,05
	IGHV3-21	**IGLV 3-1**	0,05

	IGHV3-21	IGLV 2-8	0,05
	IGHV3-21	IGLV 6-57	0,05
	IGHV3-21	IGLV 3-25	0,05
	IGHV3-21	IGLV 7-46	0,05
	IGHV3-15	**IGLV 2-14**	0,05
	IGHV3-15	**IGLV 1-47**	0,05
	IGHV3-15	**IGLV 2-23**	0,05
	IGHV3-15	**IGLV 3-21**	0,05
	IGHV3-15	IGLV 6-57	0,05
	IGHV3-15	IGLV 3-25	0,05
	IGHV3-15	IGLV 2-18	0,05
	IGHV3-15	IGLV 3-22	0,05
	IGHV4-31	IGLV 1-44	0,05
	IGHV4-31	**IGLV 2-11**	0,05
	IGHV4-31	**IGLV 3-1**	0,05
	IGHV4-31	IGLV 4-69	0,05
	IGHV4-31	IGLV 7-43	0,05
	IGHV1-2	**IGLV 3-21**	0,05
	IGHV1-2	**IGLV 2-11**	0,05
	IGHV1-2	IGLV 3-27	0,05
	IGHV3-33	**IGLV 1-40**	0,05
	IGHV3-33	IGLV 1-44	0,05
	IGHV3-33	**IGLV 1-51**	0,05
	IGHV3-33	**IGLV 2-11**	0,05
	IGHV3-33	**IGLV 3-1**	0,05
	IGHV3-33	IGLV 4-69	0,05
	IGHV3-33	IGLV 3-27	0,05
	IGHV3-33	IGLV 9-49	0,05
	IGHV3-33	IGLV 3-9	0,05
	IGHV3-53	**IGLV 1-51**	0,05
	IGHV3-53	**IGLV 1-47**	0,05
	IGHV3-53	**IGLV 2-23**	0,05
	IGHV3-53	**IGLV 2-11**	0,05
	IGHV3-53	**IGLV 3-1**	0,05
	IGHV3-53	IGLV 2-8	0,05
	IGHV3-53	IGLV 7-46	0,05
	IGHV3-11	**IGLV 1-40**	0,05
	IGHV3-11	**IGLV 1-51**	0,05
	IGHV3-11	**IGLV 1-47**	0,05
	IGHV3-11	IGLV 2-8	0,05
	IGHV3-11	IGLV 3-25	0,05
	IGHV3-11	IGLV 7-46	0,05
	IGHV3-11	IGLV 9-49	0,05
	IGHV3-11	IGLV 8-61	0,05
	IGHV3-9	IGLV 1-40	0,05
	IGHV3-9	IGLV 1-51	0,05
	IGHV3-9	IGLV 4-69	0,05
	IGHV3-9	IGLV 4-60	0,05

	IGHV3-74	**IGLV 1-47**	0,05
	IGHV3-74	**IGLV 2-11**	0,05
	IGHV3-74	**IGLV 3-1**	0,05
	IGHV3-74	IGLV 2-8	0,05
	IGHV3-74	IGLV 7-43	0,05
	IGHV3-74	IGLV 7-46	0,05
	IGHV4-4	**IGLV 2-11**	0,05
	IGHV4-4	**IGLV 3-1**	0,05
	IGHV4-4	IGLV 3-25	0,05
	IGHV4-4	IGLV 9-49	0,05
	IGHV1-46	**IGLV 1-40**	0,05
	IGHV1-46	**IGLV 1-47**	0,05
	IGHV1-46	**IGLV 2-23**	0,05
	IGHV1-46	**IGLV 3-21**	0,05
	IGHV1-46	IGLV 6-57	0,05
	IGHV4-61	IGLV 2-23	0,05
	IGHV4-61	IGLV 3-21	0,05
	IGHV4-61	IGLV 3-1	0,05
	IGHV4-61	IGLV 7-43	0,05
	IGHV1-8	IGLV 1-51	0,05
	IGHV1-8	IGLV 2-11	0,05
	IGHV1-8	IGLV 2-8	0,05
	IGHV1-8	IGLV 9-49	0,05
	IGHV1-24	IGLV 2-14	0,05
	IGHV1-24	IGLV 1-40	0,05
	IGHV1-24	IGLV 1-44	0,05
	IGHV1-24	IGLV 3-21	0,05
	IGHV1-24	IGLV 2-11	0,05
	IGHV1-3	IGLV 1-40	0,05
	IGHV3-49	IGLV 2-14	0,05
	IGHV3-49	IGLV 1-40	0,05
	IGHV3-49	IGLV 2-23	0,05
	IGHV3-49	IGLV 2-8	0,05
	IGHV4-28	IGLV 2-14	0,05
	IGHV3-43	IGLV 2-14	0,05
	IGHV3-43	IGLV 2-11	0,05
	IGHV3-43	IGLV 3-1	0,05
	IGHV3-43	IGLV 1-36	0,05
	IGHV3-43	IGLV 9-49	0,05
	IGHV3-64	IGLV 2-14	0,05
	IGHV3-64	IGLV 7-43	0,05
	IGHV7-81	IGLV 1-40	0,05
	IGHV3-13	IGLV 1-40	0,05
	IGHV3-13	IGLV 1-47	0,05
	IGHV3-72	IGLV 1-51	0,05
	IGHV3-72	IGLV 4-69	0,05
	IGHV3-73	**IGLV 1-40**	0,05
	IGHV3-73	**IGLV 1-51**	0,05

	IGHV3-73	**IGLV 1-47**	0,05
	IGHV3-73	**IGLV 2-11**	0,05
	IGHV3-73	IGLV 6-57	0,05
	IGHV1-58	IGLV 2-14	0,05
	IGHV3-66	IGLV 1-44	0,05
	IGHV3-66	IGLV 1-47	0,05
	IGHV3-66	IGLV 3-25	0,05
	IGHV4-30.2	IGLV 3-21	0,05
	IGHV7-4.1	IGLV 1-51	0,05
	IGHV3-20	IGLV 2-14	0,05

Example 5: Generation of germline genes for functional analysis

As a next step, the VH, Vλ, and Vκ germline genes selected for combination and subsequent testing, as shown in Table 5, were sent to Geneart (Regensburg, Germany) for codon optimization respective to *E. coli* expression (neutral to mammalian expression with no rare human codons), gene optimization to remove potential inhibitory or splice motifs and synthesis.

The germline protein sequences of each of the VH, Vλ, and Vκ germline genes are shown in Figs. 6-8. Each germline gene sequence was synthesized as follows:

a) for VH: leader sequence (modified phoA signal sequence incorporating a NheI restriction site as shown in Table 1); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BssHII restriction site (GCGCGC) as shown in Fig. 1); CDR-H3 (WGGDGFYAMDY) (SEQ ID NO: 1) of the 4D5 antibody as used in Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the JH4 FR4 (incorporating a XhoI (CTCGAG) restriction site as shown in Fig. 1);

b) for Vk: leader sequence (modified ompA signal sequence incorporating the NdeI restriction site as shown in Table 2); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BbsI restriction site (GAAGAC) as shown in Fig. 1), kappa-like CDR-L3 (QQHYTTPPT) (SEQ ID NO: 2) according to Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the Jk1 FR4 (incorporating a KpnI/Acc65I RE site (GGTACC) as shown in Fig. 1);

c) for Vλ: leader sequence (modified ompA signal sequence incorporating the NdeI restriction site as shown inTable 2); germline FR1, CDR1, FR2, CDR2 and FR3 (incorporating a BbsI restriction site (GAAGAC) as shown in Fig. 1), lambda-like CDR-L3 (QSYDSSLSGVV) (SEQ ID NO: 3) according to Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553; and the Jl2/3 FR4 (incorporating a KpnI/Acc65I RE site (GGTACC) as shown in Fig. 1).

Example 6: Functional testing of Germline protein pairs representative of the human immune repertoire

The 400 Germline protein pairs were then inserted into phage display, E.Coli and mamallan expression vectors either in Fab or human IgG1 format and then tested for the following properties: a) relative display after phage production and phage ELISA in Fab format; b) relative Fab expression yield after Fab production in E. coli, E. coli cell lysis and ELISA detection of produced Fab; c) temperature stability of Fab after Fab production in E. coli, E. coli cell lysis and ELISA detection of non-denatured Fab after incubation at increased temperatures; d) bovine/mouse serum stability of Fab from E. coli lysates by ELISA detection of non-denatured Fab after incubation in bovine/mouse serum; e) relative human IgG1 expression yield after IgG1 production in mammalian cells and ELISA detection of secreted IgG1 from cell culture supernatants; and f) bovine serum stability of human IgG1 by ELISA detection of non-denatured IgG after incubation in bovine/mouse serum.

Example 6.1: Generation of Fab pool displayed on phage for functional characterization

The antibody or antibody fragments synthesized in Example 5, shown in Table 5, were cloned into the tricistronic Fab display vector pJPd1 (Fig. 9) for functional testing. Fab pools were generated that contained combinations of each of the master genes, the 20 VH, combined with the 8 Vλ and 12 Vκ, yielding the 400 combinations shown in Table 6.

Phage comprising the above gene pairs were produced in a small scale using 96 well plates. A master plate was generated by filling each of the wells with 2xYT/CAM/TET/Gluc medium and inoculating with clones from the 400 VH/VL combinations wherein pMORPH30_ Vk3-11_AQA / VH3-23_TKA or pMORPH30_Vk3-11_AYA / VH3-23_VLA (pMORPH30 is shown in Fig. 12) were used as a control. The plates were incubated overnight at 37°C while shaking. The master plates were stored in a final concentration of 15% glycerol, and frozen at -80°C.

Additional 96 well plates were produced for phage production using 2xYT/CAM/TET/Gluc as medium and inoculated with clones from the master plates described above. The plates were incubated at 37°C for ~2-4h while shaking at 400 rpm, until an OD600nm of ~0,5 was reached.

The plates were infected with 5 µl helper phage per well (Hyperphage; PROGEN; 1 x 1012 pfu/ml). The plates were incubated at 37°C for 45 min without shaking and then for 60 min while shaking at 400 rpm. The bacteria were spun down at 2200g for 5 min at 4°C.

The helper phage containing supernatants were discarded and the infected E. coli pellets were re-suspended with 2xYT/Cam/TET/Kan/ IPTG without glucose. The re-suspended

pellets were transferred into a new 96 deep well plate pre-filled with 2xYT/Cm/TET/Kan/ IPTG. The plates were incubated overnight at 22°C, while shaking. The phage containing supernatants were harvested by spinning down and discarding E. coli cells and debris.

Example 6.2: Evaluation of Fab phage display ranking using ELISA

The phage supernatants prepared as described in Example 6.1 were used for Fab phage display ranking in phage ELISAs. Display of the Fab fragments was evaluated in a phage ELISA using two different capture antibodies:

(1) The anti-M13 antibody (Amersham #27-9420-01) was used for capture of phage particles via the major coat protein g8p; therefore, phage titer can be determined.

(2) An anti-Fd antibody (The Binding Site #PC075) was used, which binds to the displayed Fab; therefore, only phage displaying Fabs comprising the master genes, are captured.

The respective capture antibodies were immobilized on black 96-well Maxisorp™ plates by dispensing 100 µl antibody solution at a concentration of 7.5 µg/ml for the anti-M13 antibody and a 1.0 µg/ml concentration for the anti-Fd antibody into different wells, sealing the plate with laminated foil and incubating overnight at 4°C. The next day, the plates were washed twice with TBST, and each well was blocked with 300 µl CTBST for 1 h at room temperature.

Both the phage supernatants and reference samples were transferred for detection as follows. The blocked ELISA plates were washed twice with TBST. 100 µl of appropriately diluted phage supernatants in CTBST was transferred from the dilution plates to the coated ELISA plates, incubated for 1 – 2 h at room temperature, and washed 5x with TBST. 100 µl / well of anti-M13 peroxidase conjugate (Amersham) diluted 1:5000 in CTBST was added, and incubated for 1 - 2 h at room temperature. The Quanta Blu (Pierce) working solution was prepared by mixing 1 part (e.g. 0.5 ml) peroxide solution with 9 parts (e.g. 4.5 ml) substrate solution and equilibrating it to room temperature for at least 30 min. The ELISA plates were washed 5x with TBST, 100 µl / well of the QuantaBlu working solution was added. The fluorescence was measured after an incubation time of ~ 2 min (excitation: 320 nm, emission: 430 nm) and subsequently at intervals of 5 min.

The evaluation of the ELISA data was completed as follows: calibration curves were created by using a HuCAL GOLD reference phage preparation (VH3 kappa + lambda) and the titers of the phage supernatants and controls were calculated. For each sample, the titer on anti-Fd was divided by the titer on anti-M13 (anti-pVIII), the resulting ratio is the relative display rate. Table 12 shows the relative display rates for most of the 400 Germline protein pairs.

Example 6.3: Screening ELISA of 400 VH/VL combinations to determine the Fab expression yield in E. coli lysates

Masterplates (MP) were inoculated by picking clones transformed by pools of VH/VL combinations in the Fab expression vector pJPx1 (shown in Fig. 10) into 2YT/Cam/1%Gluc medium per well. These plates were incubated at 37°C over night while shaking. Expression plates (EP) were inoculated with 2.5 µl of the cultures from MPs into 2YT/Cam/0,1%Glucose per well. Controls (see Table 8) were inoculated from glycerol stocks. These plates were incubated for 6 hours at 37°C and shaking, then Fab expression was induced by adding IPTG and incubated at 22°C over night while shaking. E. coli cell lysates were produced by adding boric/acid/EDTA/lysozyme-buffer to the EPs (1 h incubation at 22°C, shaking), and bacterial lysates were subsequently blocked with 12,5% MPBST, shaking at least for 30 min at room temperature. E. coli lysates from expression plates were diluted appropriately in 0.5% MPBS and used in the following assay.

Table 7 shows the unlabeled coating antibodies and AP-labeled detection antibodies which were used.

Table 7:

	MOR Name	Label	Host	Antibody	Company	Number	Concentration	Dilution	Lot
Coating Ab	15	unlabeled	sheep	anti-Human IgG (Fd)	Binding Site	pc075	12.1 mg/ml	1:1000	236366, Exp 2009/10
detection Ab	AP27	AP	mouse	anti-FlagM2	Sigma	A9469	1.1 mg/ml	1:5000	048K6143, new lot

Table 8 describes the controls used.

Table 8:

#	Construct name
3	pMx11_FH VH1-69 VLA_Vl1-40 AYA
5	pMx11_FH VH3-23 VLA_Vk3-11 AYA
empty BEL	pMx9_APStuffer_FHClone1 (not containing Fab molecules!)

The screening ELISA comprised the following steps: Coating 384 wells of a MaxiSorp plate with anti-human IgG Fd specific antibodies diluted in PBS, and incubating over night at 4 °C. The next day, the plates were washed 2 x with PBST and blocked by adding (5 % Milkpowder in PBS) to each well and incubating for 1-2 h at RT, while shaking. Then the plates were washed again with PBST, and preblocked E. coli -lysates, diluted in 0,5 % MPBS, were added and incubated for 1 h while shaking at RT. Also the controls #3 and #5, were added. The plates were then washed with PBST and the AP-labeled detection antibody was diluted in 0,5% MPBS. The diluted detection antibody was added and then incubated for 1 h at RT while

shaking gently. The signal was identified by the following: washing the wells with TBST and adding 20 µl of AttoPhos (1:5 diluted in ddH2O), and reading at 5 min and 7-8 min using Tecan (infiniTe F200), program PrimeScreen.

Relative Fab expression yields are calculated by dividing the ELISA signal of the respective VH/VL pair through the ELISA signal of the reference Fab pMx11_FH VH1-69 VLA_VI1-40 AYA. Thereby equally high ELISA signals result in a relative Fab expression yield of 1. The reference Fab is expressed in a pMORPHX11 plasmids (shown in Fig. 11) comprising a) the modified phoA heavy chain signal sequence comprising the C-terminal NheI restriction site; b) the modified ompA light chain signal sequence comprising the C-terminal NdeI restriction site; c) the variable heavy germline protein sequences of the VH1-69* 01 germline gene as shown in Figure 6A, d) the variable light germline protein sequences of the IGLV1-40 germline gene as shown in Figure 8A; e) incorporating the CDR-H3 (WGGDGFYAMDY) (SEQ ID NO: 1) of the hu4D5-8 antibody, and the JH4 germline protein sequence for heavy chain FR4; f) incorporating the CDR-L3 region (QSYDSSLSGVV) (SEQ ID NO: 3) and the JI2/3 germline protein sequence for light chain FR4. The hu4D5-8 is described in Carter P. et al. (1992) "Humanization of an anti-p185Her2 antibody for human cancer therapy" Proc.Natl. Acad. Sci. USA 89, 4285-4289) and Ewert S. et al., J. Mol. Biol. (2003) 325, 531–553. All genes were generated at Geneart (Regensburg, Germany). The results are shown in Table 12.

Example 6.4: Screening ELISA of 400 VH/VL combinations to determine the temperature stability of Fab in BEL lysates

Expression plates were generated as in Example 6.3. Diluted E.coli lysates from expression plates were incubated at different temperatures for 45 minutes and used in the following assay. Table 9 shows the unlabeled coating antibodies and AP-labeled detection antibodies which were used.

Table 9:

	MOR Name	Label	Host	Antibody	Company	Number	Concen-tration	Dilution	Lot
coating Ab	57	unlabeled	Mouse	monoclonal Anti poly Histidine Antibody IgG1 (anti 6x-Histidine); polypeptides containing a polyhistidine tag	R&D Systems	MAB050	500 µg/ml	1:250	AEJ1708111
detection Ab	AP30	AP	goat	anti-human kappa light chains	Sigma	A3813	2.3 mg/ml	1:2300	018K6069
detection Ab	AP5	AP	goat	anti-human lambda light chains	Sigma	A2904	0.8 mg/ml	1:800	096K6030

The screening ELISA comprised the following steps: 384 wells of a MaxiSorp plate were coated with coating antibody (see table above) diluted in PBS. The plates were incubated over night at 4 °C. The next day, the plates were washed with PBST and blocked by adding 5 %

MPBS to each well and incubated for 1-2 h at RT while shaking. Then the diluted E.coli lysates from the expression plates were distributed into four 96 well PCR-plates (each about 40 µl) and exposed to different temperatures (4°C (on ice), 60°C, 70°C, 80°C and then on ice) in a PCR-Cycler, each temperature for 45 min. The blocked 384 well plates were washed with PBST, then the pre-incubated Fab lysates, were added to the plates. The plates were then incubated 1 h at RT while shaking. The plates were washed with PBST, the AP-labeled detection antibodies were diluted in 0.5% MPBS. 20 µl/well of the diluted detection antibodies were added and incubated for 1 h at RT while shaking gently. The signal was identified by the following: washing the wells with TBST and adding AttoPhos (1:5 diluted in ddH2O) to all wells. The signal was read at different timepoints (5 min to 10 min) using Tecan (infiniTe F200), program PimeScreen. The results are shown in Table 12.

Example 6.5: Screening ELISA of 400 VH/VL combinations to determine the serum stabilty of Fab in E.coli lysates

Expression plates were generated as in Example 6.3. The Fab containing E.coli lysates were diluted and incubated in bovine and mouse serum using the following steps: E.coli lysates from the expression plates were diluted in 50 % serum (total volume of 100 µl), 1:1000 Cam was added to prevent growth of bacteria, and the lysates were split into two 96 well plates and both plates were frozen. The first plate was thawed and incubated at 37 °C for 12-13 days. The second plate was stored at -80°C until performing the ELISA (0 days incubation at 37°C).
Table 10 shows the unlabeled coating antibodies and AP-labeled detection antibodies which were used.

Table 10:

	MOR Name	Label	Host	Antibody	Company	Number	Concen-tration	Dilution	Lot
coating Ab	36	Fab	Goat	anti-Human IgG (H+L)	Jackson Immuno Research	109-006-088	1.3 mg/ml	1:1000	80299
detection Ab	AP30	AP	goat	anti-human kappa light chains	Sigma	A3813	2.3 mg/ml	1:2300	018K6069
detection Ab	AP5	AP	Goat	anti-Human lambda-light chain; bound + free	Sigma	A2904	0.8 mg/ml	1:800	096K6030

On day 11 or 12, the 384 wells of a MaxiSorp plate were coated with 20 µl coating antibody diluted in PBS. The plates were incubated over night at 4 °C. The following day, the plates were washed with PBST and blocked by adding 5 % MPBS to each well and incubating for 1-2 h at RT while shaking. Then the blocked 384 well plates were washed with PBST. E.coli lysates in serum from the -80°C and 37°C samples were transferred to the coated ELISA plates and incubated for 1 hour at RT while shaking. The plates were washed with PBST, and the AP-

labeled detection antibodies were diluted in 0,5% MPBS. AP-labeled detection antibody was added and the plate was incubated for 1 h at RT while shaking. The signal was identified by the following: washing the wells with TBST and adding AttoPhos (1:5 diluted in ddH2O) to all wells. The signal was read at different timepoints (5 min to 10 min) using Tecan (infiniTe F200), program PrimeScreen. The results of the bovine serum stability testing are shown in Fig. 19. The results of the mouse serum stability testing are shown in Table 12.

Example 7: Generation of human IgG1 for evaluation of biophysical properties

For generation of the 400 IgG1 germline protein pairs, the 20 variable region heavy chain genes were sub-cloned into the human IgG1 expression vector pJP_hIgG1f shown in Fig. 13. In parallel the 12 variable region kappa genes were sub-cloned into the mammalian kappa light chain expression vector pJP_hIgkappa shown in Fig. 14 and the 8 variable region lambda genes were sub-cloned into the mammalian lambda light chain expression vector pJP_hIglambda2 shown in Fig. 15.

By co-transfection of each, a heavy chain and a light chain expression plasmid for all 400 VH/VL pairs can be produced separately by only cloning 40 expression constructs. Thus HEK.EBNA cells were co-transfected with all 20 heavy chain constructs and all 20 of the light chain expression constructs. Human IgG1 was harvested or detected several days post transfection from the cell culture supernatants.

Example 7.1: IgG1 expression ranking

One of the criteria for the selection of the VH/VL pairings to be included in a collection is the level of expression of the 400 different VH/VL pairings in the IgG1 format. The expression level of each VH/VL pairing in human IgG1 format was assessed by sandwich ELISA. Therefore, HEK.EBNA cells were transfected with all 400 VH/VL combinations in human IgG1 format and expressed in small scale. The cell culture supernatants were harvested after few days and IgG levels assessed.

The following procedure was performed. 384-well MaxiSorpTM plates were coated with Fcγ-pan R10Z8E9 mouse anti-human IgG at 2.5 µg/ml in PBS. The plates were incubated overnight at 4°C. The plates were washed with PBST. The plates were blocked with 5% BSA or 1x Chemiblocker in PBST and incubated for 1h at room temperature while shaking and again washed with PBST. The IgG expression supernatants were diluted in 2.5% BSA-PBST and the diluted samples were added to the blocked and washed ELISA plate. The following controls were used: empty supernatant and supernatants with a low expressing antibody, moderate

expressing antibody and a high expressing antibody. The plates were incubated for 2 h at room temperature while shaking. The plates were then washed with TBST. Appropriately diluted Fcγ-pan R10Z8E9 mouse anti-human IgG Biotin conjugate in 1% BSA-TBST was added. The plates were incubated for 1h at room temperature. The plates were washed with TBST. Streptavidin-AP diluted 1:2000 in 0.5% BSA-TBST was added and the plates were incubated for 1h at room temperature while shaking. The plates were washed with TBST. AttoPhosTM fluorescence substrate (prepared according to manufacturer's instructions) diluted in TBST directly before use was added. After 5 and 10 min, the fluorescence was measured via Tecan microplate reader.

Relative IgG1 expression yields were calculated by dividing the ELISA signal of the respective VH/VL pair through the ELISA signal of the reference IgG1 MOR03080 (shown in Table 11). Thereby equally high ELISA signals result in a relative IgG1 expression yield level of 1.

Table 11

The amino acid sequence of MOR03080 is as follows:

03080 Variable heavy chain with CDRs in bold:

(1) QVQLVESGGGLVQPGGSLRLSCAAS**GFTFSSYGMH**WVRQAPGKGLEWVS**N**

(51) **IYSDGSNTFY ADSVKG**RFTISRDNSKNTLYLQMNSLRAEDTAVYYCAR**NM**

(101) **YRWPFHYFFDY**WGQGTLVTVSS (SEQ ID NO: 61)

03080 Variable light chain with CDRs in bold

(1) DIELTQPPSV SVAPGQTARISC**SGDNIGNKYVS**WYQQKPGQAPVVVIY**GD**

(51) **NNRPS**GIPERFSGSNSGNTATLTISGTQAEDEADYYC**SSYDSSYF**VFGGG

(101) TKLTVLGQ (SEQ ID NO: 62)

The results are shown in Table 12. The sequences of the Fc portion are shown in Figures 48, 50-51.

Example 7.2: IgG1 serum stability ranking

One of the criteria for the selection of the variable heavy and variable light chain pairings to be included in a collection is the serum stability of the 400 different variable heavy and variable light chain pairings in IgG1 format. Serum stability of each IgG antibody supernatant was assessed by incubation in 50% mouse serum for 14 days and subsequent sandwich ELISA with mouse anti-human IgG (CH2) clone R10Z8E9. Again all 400 VH/VL combinations in human IgG1 format were transfected into HEK.EBNA cells and expressed in small scale. The cell

culture supernatants were harvested after few days and the IgGs in the supernatant tested for serum stability.

The following procedure was performed. 384-well MaxiSorpTM plate were coated with Fcγ-pan R10Z8E9 mouse anti-human IgG at 2.5 µg/ml in PBS. The plates were incubated overnight at 4°C. The plates were washed with PBST and then blocked with 5% BSA-PBST or 1x Chemiblocker for 1h at room temperature while shaking. The plates were washed with PBST. The IgG1 containing cell culture supernatants were diluted a) in 2.5% BSA-PBST and b) in 50% mouse serum and incubated at 37°C for at least 14 days and these samples were added to the blocked and washed ELISA plate. The following controls were used: empty supernatant and supernatants a low expressing antibody, a moderate expressing antibody, and a high expressing antibody. The plates were incubated for 2h at room temperature while shaking. The plates were washed with TBST. Fcγ-pan R10Z8E9 mouse anti-human IgG Biotin conjugate diluted to 0.8 µg/ml in 1% BSA-TBST was added. The plates were incubated for 1h at room temperature. The plates were washed with TBST. Streptavidin-AP diluted 1:2000 in 0.5% BSA-TBST was added. The plates were incubated for 1h at room temperature while shaking. The plates were washed with TBST. AttoPhosTM fluorescence substrate (prepared according to manufacturer's instructions) diluted 1:5 in TBST directly before use was added. After 5 and 10 min, the fluorescence was measured via Tecan microplate reader. The results are shown in Table 12.

Example 8: Selection of the VH/VL pairs with favorable bio-physical properties for incorporation into collection

Once the 400 germline protein pairs were tested for the following properties: a) relative display after phage production and phage ELISA in Fab format; b) relative Fab expression yield after Fab production in E. coli, E. coli cell lysis and ELISA detection of produced Fab; c) temperature stability of Fab after Fab production in E. coli, E. coli cell lysis and ELISA detection of non-denatured Fab after incubation at increased temperatures; d) bovine/mouse serum stability of Fab from E. coli lysates by ELISA detection of non-denatured Fab after incubation in bovine/mouse serum; e) relative human IgG1 expression yield after IgG1 production in mammalian cells and ELISA detection of secreted IgG1 from cell culture supernatants; and f) bovine serum stability of human IgG1 by ELISA detection of non-denatured IgG1 after incubation in bovine/mouse serum; then the next step was to select which VH/VL germline pairs were to be incorporated into the collection. The results of the functional testing for each VH/VL germline protein pairs are shown Table 12.

Table 12: Compilation of functional data for each of the 400 Germline protein pairs

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
1	hVH_1_2	hVK_1_05	0.1	0.0	bg	U	S	10	0.0	bg
2	hVH_1_2	hVK_1_06	0.1	0.2	60	S	S	42	0.0	bg
3	hVH_1_2	hVK_1_09	0.0	0.0	bg	U	S	11	0.0	bg
4	hVH_1_2	hVK_1_12	0.0	0.0	bg	S	S	20	0.0	bg
5	hVH_1_2	hVK_1_16	0.1	0.0	bg	S	S	20	0.0	bg
6	hVH_1_2	hVK_1_17	0.0	0.0	bg	S	S	21	0.0	bg
7	hVH_1_2	hVK_1_27	0.0	0.1	bg	S	S	22	0.0	bg
8	hVH_1_2	hVK_1_39	0.0	0.0	bg	S	S	21	0.0	bg
9	hVH_1_2	hVK_2_30		0.0	bg	S	S	20	0.0	bg
10	hVH_1_2	hVK_3_11	0.0	0.0	bg	S	S	20	0.0	bg
11	hVH_1_2	hVK_3_15	0.0	0.0	bg	U	S	10	0.0	bg
12	hVH_1_2	hVK_3_20		0.0	bg	S	S	21	0.0	bg
13	hVH_1_2	hVL_1-40						0	0.3	bg
14	hVH_1_2	hVL_1-47	0.0	0.0	4	U	U	2	0.0	bg
15	hVH_1_2	hVL_1-51	0.0	0.0	4	U	U	0	0.4	bg
16	hVH_1_2	hVL_2-11	0.1	0.0	4	S	S	22	0.3	bg
17	hVH_1_2	hVL_2-14	0.1	0.0	4	U	U	0	0.1	bg
18	hVH_1_2	hVL_2-23	0.0	0.0	4	U	U	0	0.0	bg
19	hVH_1_2	hVL_3-1	0.4	0.0	4	U	U	1	0.0	bg
20	hVH_1_2	hVL_3-21	0.0	0.0	4	U	U	0	0.0	bg
21	**hVH_1_18**	**hVK_1_05**	**2.0**	**0.4**	**60**	**S**	**S**	**54**	**0.4**	**S**
22	hVH_1_18	hVK_1_06	0.6	0.5	60	S	S	56	0.2	S
23	hVH_1_18	hVK_1_09						0	0.1	S
24	hVH_1_18	hVK_1_12	1.6	0.5	60	S	S	56	0.1	bg
25	hVH_1_18	hVK_1_16	2.0					3	0.2	S
26	hVH_1_18	hVK_1_17		0.5		S	S	38	0.3	S
27	hVH_1_18	hVK_1_27	1.2	0.4	70	S	S	62	0.5	S
28	hVH_1_18	hVK_1_39	3.7	0.3	60	S	S	53	0.1	S
29	hVH_1_18	hVK_2_30	1.9	0.5	60	S	S	56	0.0	S
30	hVH_1_18	hVK_3_11		0.6	60	S	S	56	0.0	S
31	hVH_1_18	hVK_3_15	2.6	0.5	70	S	S	67	0.3	S
32	hVH_1_18	hVK_3_20	2.2	0.9	60	S	S	72	0.0	S
33	hVH_1_18	hVL_1-40	2.4					4	0.5	S
34	hVH_1_18	hVL_1-47		0.8	60	S	S	66	0.4	U
35	hVH_1_18	hVL_1-51						0	0.5	S
36	hVH_1_18	hVL_2-11	1.9					3	0.5	U
37	hVH_1_18	hVL_2-14	2.5	0.6	60	S	S	64	0.5	U
38	**hVH_1_18**	**hVL_2-23**	**4.3**	**0.7**	**60**	**S**	**S**	**70**	**0.4**	**S**
39	hVH_1_18	hVL_3-1	4.4	0.6	60	S	S	65	0.2	U
40	hVH_1_18	hVL_3-21	3.4	0.6	60	S	S	64	0.2	S
41	hVH_1_46	hVK_1_05		0.4	60	S	S	51	0.9	S
42	hVH_1_46	hVK_1_06						0	0.9	S
43	**hVH_1_46**	**hVK_1_09**	**3.0**	**0.6**	**60**	**S**	**S**	**63**	**0.4**	**S**
44	hVH_1_46	hVK_1_12		0.5	60	S	S	55	0.2	S
45	hVH_1_46	hVK_1_16	1.3	0.6	60	S	S	61	0.3	S
46	hVH_1_46	hVK_1_17	1.3					2	0.5	S
47	hVH_1_46	hVK_1_27						0	0.6	S
48	**hVH_1_46**	**hVK_1_39**	**2.5**	**0.4**	**60**	**S**	**S**	**55**	**0.5**	**S**
49	hVH_1_46	hVK_2_30		0.2	4	U	S	16	0.0	S

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
50	hVH_1_46	hVK_3_11						0	0.1	S
51	**hVH_1_46**	**hVK_3_15**	**3.0**	**0.7**	**60**	**S**	**S**	**68**	**0.4**	**S**
52	hVH_1_46	hVK_3_20						0	0.1	S
53	hVH_1_46	hVL_1-40		1.0	60	S	S	73	0.9	S
54	hVH_1_46	hVL_1-47						0	0.6	U
55	hVH_1_46	hVL_1-51	5.7					10	0.3	S
56	hVH_1_46	hVL_2-11	1.6					3	0.3	S
57	hVH_1_46	hVL_2-14						0	0.3	U
58	hVH_1_46	hVL_2-23	2.7	1.0	60	S	S	79	0.3	S
59	hVH_1_46	hVL_3-1	4.3					7	0.4	S
60	hVH_1_46	hVL_3-21	5.2					9	0.3	S
61	**hVH_1_69*01**	**hVK_1_05**	**2.1**	**0.5**	**60**	**S**	**S**	**59**	**0.9**	**S**
62	hVH_1_69*01	hVK_1_06	2.9					5	0.5	S
63	hVH_1_69*01	hVK_1_09		0.3	60	S	U	37	0.4	S
64	hVH_1_69*01	hVK_1_12	2.1	0.4	60	S	S	53	0.3	S
65	hVH_1_69*01	hVK_1_16	1.2					2	0.4	S
66	hVH_1_69*01	hVK_1_17	0.9	0.3	4	S	S	31	0.3	S
67	hVH_1_69*01	hVK_1_27	0.2	0.3	70	S	S	56	0.4	S
68	hVH_1_69*01	hVK_1_39	3.5	0.1	4	S	S	31	0.4	U
69	hVH_1_69*01	hVK_2_30						0	0.0	S
70	hVH_1_69*01	hVK_3_11		0.7	60	S	S	60	0.0	S
71	hVH_1_69*01	hVK_3_15	1.6	0.5	70	S	S	66	0.5	S
72	hVH_1_69*01	hVK_3_20		0.5	60	S	S	54	0.0	S
73	hVH_1_69*01	hVL_1-40		1.0	60	S	S	72	0.2	S
74	hVH_1_69*01	hVL_1-47						0	0.2	U
75	hVH_1_69*01	hVL_1-51		0.8	60	S	S	64	0.3	S
76	hVH_1_69*01	hVL_2-11	0.8	0.7	60	S	S	65	0.2	S
77	hVH_1_69*01	hVL_2-14		0.8	60	S	S	64	0.3	U
78	hVH_1_69*01	hVL_2-23	1.8					3	0.3	S
79	hVH_1_69*01	hVL_3-1	3.4	0.7		S	S	52	0.2	S
80	hVH_1_69*01	hVL_3-21	4.6	0.7	60	S	S	71	0.1	S
81	hVH_3_07	hVK_1_05		0.7	60	S	S	63	0.9	U
82	hVH_3_07	hVK_1_06		0.9	60	S	S	69	1.3	S
83	**hVH_3_07**	**hVK_1_09**	**6.7**	**0.4**	**60**	**S**	**S**	**50**	**1.5**	**S**
84	**hVH_3_07**	**hVK_1_12**	**10.6**	**0.9**	**70**	**S**	**S**	**97**	**0.9**	**S**
85	hVH_3_07	hVK_1_16	7.0					12	1.5	S
86	hVH_3_07	hVK_1_17	10.5	0.5	4	S	S	40	0.9	S
87	**hVH_3_07**	**hVK_1_27**	**14.5**	**0.5**	**70**	**S**	**S**	**87**	**1.8**	**S**
88	hVH_3_07	hVK_1_39	27.3	0.3	60	U	S	85	1.2	S
89	hVH_3_07	hVK_2_30	13.0					0	0.3	S
90	hVH_3_07	hVK_3_11						0	0.4	S
91	**hVH_3_07**	**hVK_3_15**	**14.5**	**0.7**	**70**	**S**	**S**	**95**	**1.8**	**S**
92	hVH_3_07	hVK_3_20						0	0.4	S
93	hVH_3_07	hVL_1-40	8.2					14	0.3	S
94	**hVH_3_07**	**hVL_1-47**	**6.3**	**1.2**	**60**	**S**	**S**	**90**	**0.8**	**U**
95	hVH_3_07	hVL_1-51		1.0	60	S	S	74	0.9	S
96	hVH_3_07	hVL_2-11						0	1.2	S
97	hVH_3_07	hVL_2-14	11.3					19	0.8	U
98	**hVH_3_07**	**hVL_2-23**	**6.9**	**0.8**	**60**	**S**	**S**	**76**	**0.7**	**S**
99	**hVH_3_07**	**hVL_3-1**	**5.0**	**0.5**	**60**	**S**	**S**	**64**	**1.2**	**S**

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
100	hVH_3_07	hVL_3-21		0.7	60	S	S	61	0.3	S
101	**hVH_3_11**	**hVK_1_05**	**5.5**	**0.5**	**60**	**S**	**S**	**65**	**0.5**	**S**
102	**hVH_3_11**	**hVK_1_06**	**4.3**	**0.6**	**60**	**S**	**S**	**64**	**1.4**	**S**
103	hVH_3_11	hVK_1_09	6.7					0	0.9	S
104	**hVH_3_11**	**hVK_1_12**	**8.2**	**0.6**	**60**	**S**	**S**	**73**	**0.9**	**S**
105	**hVH_3_11**	**hVK_1_16**	**10.3**	**0.6**	**60**	**S**	**U**	**61**	**1.2**	**S**
106	hVH_3_11	hVK_1_17						0	0.9	S
107	hVH_3_11	hVK_1_27	6.0					0	1.7	S
108	hVH_3_11	hVK_1_39	29.0					50	1.8	S
109	hVH_3_11	hVK_2_30		0.4	4	S	S	34	1.1	U
110	hVH_3_11	hVK_3_11	0.0					0	0.6	S
111	**hVH_3_11**	**hVK_3_15**	**4.6**	**0.7**	**60**	**S**	**S**	**68**	**1.6**	**S**
112	hVH_3_11	hVK_3_20						0	0.2	S
113	hVH_3_11	hVL_1-40	12.4					21	0.3	S
114	**hVH_3_11**	**hVL_1-47**	**8.1**	**0.8**	**60**	**S**	**S**	**80**	**1.3**	**U**
115	hVH_3_11	hVL_1-51		1.1	60	S	S	77	1.9	S
116	hVH_3_11	hVL_2-11	8.4					14	1.1	S
117	hVH_3_11	hVL_2-14	6.4	0.9	60	S	S	81	0.4	U
118	**hVH_3_11**	**hVL_2-23**	**8.9**	**1.0**	**60**	**S**	**S**	**88**	**0.4**	**S**
119	hVH_3_11	hVL_3-1		0.5	60	S	S	53	1.6	S
120	hVH_3_11	hVL_3-21	9.8					17	0.3	S
121	**hVH_3_15**	**hVK_1_05**	**8.1**	**0.5**	**60**	**S**	**S**	**68**	**0.4**	**S**
122	**hVH_3_15**	**hVK_1_06**	**11.7**	**0.6**	**60**	**S**	**S**	**79**	**0.8**	**S**
123	**hVH_3_15**	**hVK_1_09**	**10.0**	**0.5**	**70**	**S**	**S**	**80**	**0.9**	**S**
124	**hVH_3_15**	**hVK_1_12**	**11.5**	**0.7**	**70**	**S**	**S**	**90**	**0.7**	**S**
125	**hVH_3_15**	**hVK_1_16**	**14.5**	**0.7**	**60**	**S**	**S**	**86**	**1.5**	**S**
126	hVH_3_15	hVK_1_17	6.4	0.6	4	U	U	30	0.8	S
127	**hVH_3_15**	**hVK_1_27**	**7.8**	**0.5**	**70**	**S**	**S**	**77**	**1.7**	**S**
128	**hVH_3_15**	**hVK_1_39**	**14.2**	**0.4**	**60**	**S**	**S**	**76**	**1.8**	**S**
129	hVH_3_15	hVK_2_30		0.3	4	S	U	23	0.6	S
130	hVH_3_15	hVK_3_11	19.4					33	0.8	S
131	**hVH_3_15**	**hVK_3_15**	**12.1**	**0.6**	**70**	**S**	**S**	**70**	**1.9**	**S**
132	hVH_3_15	hVK_3_20	8.9					0	0.5	S
133	**hVH_3_15**	**hVL_1-40**	**16.7**	**0.9**	**60**	**S**	**S**	**98**	**0.1**	**S**
134	hVH_3_15	hVL_1-47	13.0	1.2	60	S	S	102	0.2	U
135	**hVH_3_15**	**hVL_1-51**	**11.0**	**1.1**	**60**	**S**	**S**	**94**	**0.9**	**S**
136	**hVH_3_15**	**hVL_2-11**	**10.5**	**0.9**	**60**	**S**	**S**	**88**	**0.8**	**S**
137	hVH_3_15	hVL_2-14	9.7	0.8	60	S	S	83	0.9	U
138	hVH_3_15	hVL_2-23	10.1					17	0.4	S
139	hVH_3_15	hVL_3-1	9.4	0.3	4	S	S	46	1.0	S
140	hVH_3_15	hVL_3-21	9.2	0.8		S	S	65	0.2	S
141	hVH_3_21	hVK_1_05	10.0					17	0.8	S
142	**hVH_3_21**	**hVK_1_06**	**16.1**	**1.0**	**60**	**S**	**S**	**99**	**0.9**	**S**
143	hVH_3_21	hVK_1_09						0	0.4	S
144	**hVH_3_21**	**hVK_1_12**	**11.3**	**0.6**	**60**	**S**	**S**	**77**	**0.5**	**S**
145	hVH_3_21	hVK_1_16		0.9	60	S	S	68	0.0	S
146	hVH_3_21	hVK_1_17	5.0					9	0.0	S
147	**hVH_3_21**	**hVK_1_27**	**8.7**	**0.6**	**60**	**S**	**S**	**78**	**0.5**	**S**
148	**hVH_3_21**	**hVK_1_39**	**11.6**	**0.5**	**60**	**S**	**S**	**54**	**0.8**	**S**
149	hVH_3_21	hVK_2_30		0.6	4	S	S	44	0.1	U

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
150	hVH_3_21	hVK_3_11						0	0.2	S
151	hVH_3_21	hVK_3_15		0.8	60	S	S	65	0.3	S
152	hVH_3_21	hVK_3_20						0	0.5	S
153	hVH_3_21	hVL_1-40		1.0	60	S	S	72	0.5	S
154	hVH_3_21	hVL_1-47	0.0	1.2	60	S	S	81	0.3	S
155	hVH_3_21	hVL_1-51						0	0.9	S
156	hVH_3_21	hVL_2-11		0.9	60	S	S	68	0.7	S
157	**hVH_3_21**	**hVL_2-14**	**6.5**	**0.9**	**60**	**S**	**S**	**81**	**1.2**	**S**
158	**hVH_3_21**	**hVL_2-23**	**8.8**	**1.0**	**60**	**S**	**S**	**90**	**0.9**	**S**
159	hVH_3_21	hVL_3-1		0.7	60	S	S	60	0.4	S
160	hVH_3_21	hVL_3-21	11.8	0.9	60	S	S	88	0.1	S
161	hVH_3_23	hVK_1_05		0.8	60	S	S	64	0.2	S
162	hVH_3_23	hVK_1_06		0.7	60	S	S	61	0.2	S
163	hVH_3_23	hVK_1_09	6.1	0.8	70	S	S	86	0.1	S
164	hVH_3_23	hVK_1_12		0.9	60	S	S	68	0.1	S
165	hVH_3_23	hVK_1_16	8.4	0.6	60	S	S	72	0.2	S
166	hVH_3_23	hVK_1_17		0.6	4	S	U	31	0.1	S
167	hVH_3_23	hVK_1_27	17.1					29	0.2	S
168	hVH_3_23	hVK_1_39	10.8					19	0.3	S
169	hVH_3_23	hVK_2_30	4.1	0.3	4	S	S	39	0.0	bg
170	hVH_3_23	hVK_3_11						0	0.0	bg
171	hVH_3_23	hVK_3_15		0.7	70	S	S	73	0.4	S
172	hVH_3_23	hVK_3_20	13.3					0	0.2	S
173	hVH_3_23	hVL_1-40						0	0.1	S
174	hVH_3_23	hVL_1-47						0	0.1	S
175	hVH_3_23	hVL_1-51	10.2	1.1	60	S	S	94	0.2	S
176	hVH_3_23	hVL_2-11	13.6					23	0.1	S
177	hVH_3_23	hVL_2-14	9.1					16	0.3	S
178	hVH_3_23	hVL_2-23	7.4	0.9	60	S	S	82	0.3	S
179	hVH_3_23	hVL_3-1	4.6	0.4	60	S	S	60	0.1	S
180	hVH_3_23	hVL_3-21	7.4	0.8	60	S	S	78	0.1	S
181	hVH_3_30	hVK_1_05						0	0.7	S
182	hVH_3_30	hVK_1_06		1.0	60	S	S	75	0.6	S
183	hVH_3_30	hVK_1_09						0	0.3	S
184	hVH_3_30	hVK_1_12	5.4	0.8	60	S	S	73	0.3	S
185	hVH_3_30	hVK_1_16		0.9	60	S	S	69	0.4	S
186	hVH_3_30	hVK_1_17						0	0.5	S
187	hVH_3_30	hVK_1_27	9.1	0.4	60	S	U	38	0.5	S
188	hVH_3_30	hVK_1_39	13.1	0.0	bg	U	U	19	1.0	S
189	hVH_3_30	hVK_2_30		0.4	4	S	U	23	0.1	bg
190	hVH_3_30	hVK_3_11		0.4	60	S	S	50	0.1	S
191	hVH_3_30	hVK_3_15		0.7	60	S	S	61	0.9	S
192	hVH_3_30	hVK_3_20		0.7	60	S	S	63	0.4	S
193	hVH_3_30	hVL_1-40						0	0.8	S
194	hVH_3_30	hVL_1-47		1.1	60	S	S	78	0.3	S
195	hVH_3_30	hVL_1-51						0	0.4	S
196	hVH_3_30	hVL_2-11		0.7	60	S	S	62	0.4	S
197	hVH_3_30	hVL_2-14		0.8	60	S	S	66	1.0	S
198	**hVH_3_30**	**hVL_2-23**	**9.5**	**1.0**	**60**	**S**	**S**	**89**	**0.5**	**S**
199	**hVH_3_30**	**hVL_3-1**	**8.8**	**0.6**	**60**	**S**	**S**	**73**	**0.5**	**S**

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
200	hVH_3_30	hVL_3-21	16.6	0.8	60	S	S	93	0.2	S
201	hVH_3_33	hVK_1_05		0.3	60	S	S	46	0.0	S
202	hVH_3_33	hVK_1_06						0	0.6	S
203	hVH_3_33	hVK_1_09		0.7	60	S	S	60	0.2	S
204	hVH_3_33	hVK_1_12		0.2	60	S	U	34	0.2	S
205	hVH_3_33	hVK_1_16						0	0.4	S
206	hVH_3_33	hVK_1_17						0	0.5	S
207	hVH_3_33	hVK_1_27		0.6	60	S	S	57	0.2	S
208	hVH_3_33	hVK_1_39						0	0.8	S
209	hVH_3_33	hVK_2_30						0	0.3	S
210	hVH_3_33	hVK_3_11						0	0.6	S
211	**hVH_3_33**	**hVK_3_15**	**12.3**	**0.6**	**60**	**S**	**S**	**77**	**0.9**	**S**
212	hVH_3_33	hVK_3_20		1.0	60	S	S	72	0.3	S
213	hVH_3_33	hVL_1-40						0	1.0	S
214	hVH_3_33	hVL_1-47		1.1	60	S	S	77	0.4	S
215	hVH_3_33	hVL_1-51						0	0.6	S
216	hVH_3_33	hVL_2-11		0.5	60	S	S	54	0.5	S
217	hVH_3_33	hVL_2-14		0.9	4	S	S	53	0.9	S
218	**hVH_3_33**	**hVL_2-23**	**17.1**	**0.5**	**60**	**S**	**S**	**82**	**0.5**	**S**
219	hVH_3_33	hVL_3-1		0.2	60	S	S	44	0.7	S
220	hVH_3_33	hVL_3-21		0.8	60	S	S	67	0.5	S
221	hVH_3_48	hVK_1_05						0	0.6	S
222	hVH_3_48	hVK_1_06						0	0.7	S
223	hVH_3_48	hVK_1_09						0	0.2	S
224	hVH_3_48	hVK_1_12						0	0.3	S
225	hVH_3_48	hVK_1_16	8.7					15	0.5	S
226	hVH_3_48	hVK_1_17						0	0.5	S
227	**hVH_3_48**	**hVK_1_27**	**8.9**	**0.7**	**60**	**S**	**S**	**74**	**0.9**	**S**
228	hVH_3_48	hVK_1_39						0	0.5	S
229	hVH_3_48	hVK_2_30						0	0.3	S
230	hVH_3_48	hVK_3_11						0	0.7	S
231	hVH_3_48	hVK_3_15	12.1					21	0.3	S
232	hVH_3_48	hVK_3_20		0.8	60	S	S	65	0.4	S
233	hVH_3_48	hVL_1-40		0.8		S	S	51	0.6	S
234	hVH_3_48	hVL_1-47	10.3					18	0.4	S
235	hVH_3_48	hVL_1-51		1.2	60	S	S	80	0.7	S
236	hVH_3_48	hVL_2-11						0	0.6	S
237	hVH_3_48	hVL_2-14						0	0.6	S
238	hVH_3_48	hVL_2-23	9.3					16	0.5	S
239	hVH_3_48	hVL_3-1	6.0	0.8		S	S	61	0.5	S
240	hVH_3_48	hVL_3-21						0	0.3	S
241	hVH_3_53	hVK_1_05	11.1	0.7	4	U	S	60	0.8	S
242	hVH_3_53	hVK_1_06		0.7	60	S	S	63	0.7	S
243	**hVH_3_53**	**hVK_1_09**	**8.3**	**0.9**	**60**	**S**	**S**	**83**	**0.4**	**S**
244	**hVH_3_53**	**hVK_1_12**	**14.8**	**0.7**	**60**	**S**	**S**	**60**	**0.2**	**S**
245	hVH_3_53	hVK_1_16	10.7	0.0	bg	bg	U	20	0.3	S
246	hVH_3_53	hVK_1_17	2.9	0.5	4	S	S	42	0.5	S
247	hVH_3_53	hVK_1_27	6.9	0.4	60	S	S	62	0.2	S
248	hVH_3_53	hVK_1_39		0.6	60	S	S	56	0.2	S
249	hVH_3_53	hVK_2_30	1.3	0.3	4	S	S	32	0.0	bg

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
250	hVH_3_53	hVK_3_11		0.8	60	S	S	64	0.3	S
251	hVH_3_53	hVK_3_15	9.6	0.7	60	S	S	63	0.5	S
252	hVH_3_53	hVK_3_20		0.3	4	S	S	32	0.3	S
253	hVH_3_53	hVL_1-40		1.1	4	S	S	60	1.1	S
254	hVH_3_53	hVL_1-47		1.1	60	S	S	79	0.2	S
255	**hVH_3_53**	**hVL_1-51**	**6.4**	**1.3**	**60**	**S**	**S**	**96**	**0.4**	**S**
256	hVH_3_53	hVL_2-11	7.2	0.8	60	S	S	78	0.3	S
257	hVH_3_53	hVL_2-14		1.0	60	S	S	75	0.8	S
258	**hVH_3_53**	**hVL_2-23**	**6.3**	**1.1**	**60**	**S**	**S**	**86**	**0.6**	**S**
259	**hVH_3_53**	**hVL_3-1**	**5.1**	**0.6**	**60**	**S**	**S**	**67**	**0.5**	**S**
260	hVH_3_53	hVL_3-21		0.8	60	S	S	66	0.5	S
261	hVH_3_73	hVK_1_05	0.4	0.2	60	S	S	45	1.1	S
262	hVH_3_73	hVK_1_06	0.3	0.2	60	S	S	45	1.0	S
263	hVH_3_73	hVK_1_09	0.3	0.1	60	S	S	39	0.9	S
264	hVH_3_73	hVK_1_12	0.3	0.1	60	S	S	38	0.5	S
265	hVH_3_73	hVK_1_16	0.3	0.2	60	S	S	44	1.1	S
266	hVH_3_73	hVK_1_17	0.1					0	1.0	S
267	hVH_3_73	hVK_1_27	3.6	0.1	4	S	S	24	0.9	S
268	hVH_3_73	hVK_1_39	0.2	0.2	4	S	S	27	0.8	S
269	hVH_3_73	hVK_2_30		0.1	bg	S	S	22	0.3	S
270	hVH_3_73	hVK_3_11	0.5					0	0.2	S
271	hVH_3_73	hVK_3_15	0.2	0.1	60	S	S	39	0.1	S
272	hVH_3_73	hVK_3_20						0	1.1	S
273	hVH_3_73	hVL_1-40		0.1	60	S	S	40	1.2	S
274	hVH_3_73	hVL_1-47	0.0	0.3	4	S	S	31	0.8	S
275	hVH_3_73	hVL_1-51	0.3	0.2	60	S	S	44	0.7	S
276	hVH_3_73	hVL_2-11	0.2	0.2	4	S	S	26	0.8	S
277	hVH_3_73	hVL_2-14						0	0.4	S
278	hVH_3_73	hVL_2-23	0.8					1	0.1	S
279	hVH_3_73	hVL_3-1	0.0	0.1	60	S	S	39	1.0	S
280	hVH_3_73	hVL_3-21	0.4	0.2	60	S	S	43	1.1	S
281	hVH_3_74	hVK_1_05	6.4					11	0.6	S
282	**hVH_3_74**	**hVK_1_06**	**9.5**	**0.9**	**60**	**S**	**S**	**86**	**1.0**	**S**
283	**hVH_3_74**	**hVK_1_09**	**8.7**	**0.6**	**60**	**S**	**S**	**74**	**0.5**	**S**
284	hVH_3_74	hVK_1_12	8.4	0.6	60	S	S	74	0.0	S
285	hVH_3_74	hVK_1_16	8.0					11	0.8	S
286	hVH_3_74	hVK_1_17		0.6	60	S	S	58	0.2	S
287	**hVH_3_74**	**hVK_1_27**	**5.0**	**0.6**	**70**	**S**	**S**	**77**	**1.1**	**S**
288	hVH_3_74	hVK_1_39	8.7					15	0.3	S
289	hVH_3_74	hVK_2_30		0.4		S	S	37	0.7	S
290	hVH_3_74	hVK_3_11						0	0.1	S
291	**hVH_3_74**	**hVK_3_15**	**10.0**	**0.8**	**70**	**S**	**S**	**94**	**1.0**	**S**
292	hVH_3_74	hVK_3_20		0.7	60	S	S	62	0.6	S
293	hVH_3_74	hVL_1-40	8.8	0.4	4	S	S	51	1.3	S
294	hVH_3_74	hVL_1-47	3.2	1.2		S	S	72	0.6	S
295	**hVH_3_74**	**hVL_1-51**	**7.1**	**1.1**	**60**	**S**	**S**	**91**	**1.2**	**S**
296	hVH_3_74	hVL_2-11		0.6	60	S	S	59	0.8	S
297	hVH_3_74	hVL_2-14	4.7					8	0.6	S
298	hVH_3_74	hVL_2-23						0	1.0	S
299	hVH_3_74	hVL_3-1	7.0	0.6	60	S	S	70	0.3	S

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
300	hVH_3_74	hVL_3-21	1.8	0.6	60	S	S	60	0.3	S
301	hVH_4_04*03	hVK_1_05		0.8	60	S	S	67	0.6	S
302	hVH_4_04*03	hVK_1_06		0.8	60	S	S	64	1.1	S
303	hVH_4_04*03	hVK_1_09	4.5	0.1	bg	S	S	30	0.6	S
304	hVH_4_04*03	hVK_1_12		0.7	60	S	S	61	0.8	S
305	hVH_4_04*03	hVK_1_16	3.2	0.2	60	S	S	48	0.4	S
306	hVH_4_04*03	hVK_1_17		0.4	4	S	S	34	0.8	S
307	hVH_4_04*03	hVK_1_27		0.4	60	S	S	48	0.9	S
308	hVH_4_04*03	hVK_1_39		0.2	bg	S	S	26	1.0	S
309	hVH_4_04*03	hVK_2_30	0.3	0.5	4	S	S	38	0.2	U
310	hVH_4_04*03	hVK_3_11		0.6	bg	S	S	43	0.3	S
311	hVH_4_04*03	hVK_3_15		0.6	60	S	S	58	1.1	S
312	hVH_4_04*03	hVK_3_20		1.1	60	S	U	65	1.1	S
313	hVH_4_04*03	hVL_1-40		1.0	60	S	S	75	0.9	S
314	hVH_4_04*03	hVL_1-47	8.3					14	0.4	S
315	hVH_4_04*03	hVL_1-51		0.9	60	S	S	71	0.6	S
316	hVH_4_04*03	hVL_2-11		1.0	60	S	S	73	0.7	S
317	hVH_4_04*03	hVL_2-14		0.7	60	S	S	63	0.4	S
318	**hVH_4_04*03**	**hVL_2-23**	**2.7**	**1.0**	**60**	**S**	**S**	**77**	**0.7**	**S**
319	**hVH_4_04*03**	**hVL_3-1**	**2.2**	**0.6**	**60**	**S**	**S**	**63**	**1.3**	**S**
320	**hVH_4_04*03**	**hVL_3-21**	**5.2**	**0.7**	**60**	**S**	**S**	**69**	**0.5**	**S**
321	hVH_4_31	hVK_1_05		0.0	bg	S	S	21	0.0	bg
322	hVH_4_31	hVK_1_06						0	0.2	bg
323	hVH_4_31	hVK_1_09		0.1	4	S	S	23	0.6	S
324	hVH_4_31	hVK_1_12		0.1	60	S	S	37	0.4	S
325	hVH_4_31	hVK_1_16		0.0	bg	S	S	20	0.0	bg
326	hVH_4_31	hVK_1_17		0.0	bg	U	bg	1	0.2	bg
327	hVH_4_31	hVK_1_27		0.0	bg	S	S	20	0.0	bg
328	hVH_4_31	hVK_1_39		0.8	60	S	S	65	0.5	S
329	hVH_4_31	hVK_2_30		0.0	bg	S	S	20	0.0	bg
330	hVH_4_31	hVK_3_11						0	0.0	bg
331	hVH_4_31	hVK_3_15		0.1	bg	S	S	24	0.1	S
332	hVH_4_31	hVK_3_20						0	0.4	S
333	hVH_4_31	hVL_1-40	0.0	0.6	60	S	S	57	0.8	S
334	hVH_4_31	hVL_1-47	0.0	0.7	60	S	S	62	0.1	S
335	hVH_4_31	hVL_1-51		0.9	60	S	S	70	0.3	S
336	hVH_4_31	hVL_2-11		0.5	60	S	S	55	0.2	S
337	hVH_4_31	hVL_2-14	0.0					0	0.5	S
338	hVH_4_31	hVL_2-23		0.0	60	S	S	37	0.3	S
339	hVH_4_31	hVL_3-1	1.4	0.3	60	S	S	50	1.3	S
340	hVH_4_31	hVL_3-21		0.4	60	S	S	50	0.4	bg
341	hVH_4_39	hVK_1_05	0.0	0.3	60	S	S	45	0.3	S
342	hVH_4_39	hVK_1_06	1.6					3	0.8	S
343	hVH_4_39	hVK_1_09		0.5	4	S	S	37	0.7	S
344	hVH_4_39	hVK_1_12						0	0.9	S
345	hVH_4_39	hVK_1_16						0	0.5	S
346	hVH_4_39	hVK_1_17	0.7	0.3	4	S	S	33	1.0	S
347	hVH_4_39	hVK_1_27						0	0.4	S
348	hVH_4_39	hVK_1_39	2.1	0.3	60	S	S	48	1.2	S
349	hVH_4_39	hVK_2_30		0.2	4	S	S	27	0.2	S

No.	VH	VL	Relative Fab Display (CysDisplay)	Relative Fab expression	Fab thermo-stability	Fab stability in mouse serum	Fab stability in bovine serum	Fab ranking-value	Relative IgG1 expression	IgG1 stability in bovine serum
350	hVH_4_39	hVK_3_11		0.3	60	S	S	48	0.2	S
351	hVH_4_39	hVK_3_15		0.6	70	S	S	68	1.0	S
352	hVH_4_39	hVK_3_20		0.6	60	S		49	1.2	S
353	hVH_4_39	hVL_1-40	0.6	0.9	70	S	S	81	1.1	S
354	hVH_4_39	hVL_1-47		0.7	70	S	S	72	0.3	S
355	hVH_4_39	hVL_1-51		0.8	60	S	S	65	0.5	S
356	hVH_4_39	hVL_2-11						0	0.3	S
357	**hVH_4_39**	**hVL_2-14**	**2.0**	**0.6**	**60**	**S**	**S**	**63**	**0.5**	**S**
358	hVH_4_39	hVL_2-23	0.9	0.7	60	S	S	62	0.4	S
359	**hVH_4_39**	**hVL_3-1**	**3.6**	**0.5**	**60**	**S**	**S**	**59**	**0.9**	**S**
360	hVH_4_39	hVL_3-21		0.6	60	S	S	57	0.6	S
361	hVH_5_51	hVK_1_05		0.5	60	S	S	52	0.4	S
362	hVH_5_51	hVK_1_06		0.5	60	S	S	54	0.9	S
363	**hVH_5_51**	**hVK_1_09**	**2.6**	**0.5**	**60**	**S**	**S**	**57**	**0.5**	**S**
364	hVH_5_51	hVK_1_12	1.8					3	0.8	S
365	hVH_5_51	hVK_1_16	1.3					2	0.5	S
366	hVH_5_51	hVK_1_17		0.3	4	S	S	32	0.6	S
367	hVH_5_51	hVK_1_27	0.4	0.2	60	S	S	43	1.0	S
368	hVH_5_51	hVK_1_39	3.7	0.3	60	S	S	51	1.2	S
369	hVH_5_51	hVK_2_30	0.9	0.2	4	S		19	0.7	S
370	hVH_5_51	hVK_3_11		1.0	60	S		62	0.6	S
371	hVH_5_51	hVK_3_15	1.9					3	1.2	S
372	hVH_5_51	hVK_3_20						0	1.1	S
373	hVH_5_51	hVL_1-40		1.0	60	S	S	72	1.3	S
374	hVH_5_51	hVL_1-47		1.0	60	S	S	73	0.8	S
375	hVH_5_51	hVL_1-51		1.1	60	S	S	77	0.5	S
376	hVH_5_51	hVL_2-11	0.0	0.7	60	S	S	63	0.3	S
377	hVH_5_51	hVL_2-14	2.1					4	0.8	S
378	**hVH_5_51**	**hVL_2-23**	**3.0**	**1.0**	**60**	**S**	**S**	**79**	**0.7**	**S**
379	**hVH_5_51**	**hVL_3-1**	**3.8**	**0.7**	**60**	**S**	**S**	**67**	**1.3**	**S**
380	hVH_5_51	hVL_3-21						0	0.7	S
381	hVH_6_1	hVK_1_05		0.7	60	S	S	62	0.0	S
382	**hVH_6_1**	**hVK_1_06**	**3.3**	**0.6**	**60**	**S**	**S**	**64**	**1.2**	**S**
383	hVH_6_1	hVK_1_09	5.9					10	1.3	S
384	hVH_6_1	hVK_1_12	1.5	0.0	bg	U	S	13	1.1	S
385	hVH_6_1	hVK_1_16						0	1.4	S
386	hVH_6_1	hVK_1_17		0.5	60	S	S	54	1.3	S
387	hVH_6_1	hVK_1_27		0.5	70	S	S	63	1.2	S
388	hVH_6_1	hVK_1_39		0.3	60	S	S	45	1.1	S
389	hVH_6_1	hVK_2_30		0.3	4	S	S	32	0.3	S
390	hVH_6_1	hVK_3_11						0	0.9	S
391	hVH_6_1	hVK_3_15		0.7	70	S	S	70	1.3	S
392	hVH_6_1	hVK_3_20		0.9	60	S	S	70	1.3	S
393	hVH_6_1	hVL_1-40	7.2					12	1.4	S
394	hVH_6_1	hVL_1-47		1.1	60	S	S	75	0.2	S
395	hVH_6_1	hVL_1-51		1.1	60	S	S	75	0.5	S
396	hVH_6_1	hVL_2-11	1.0	1.0	60	S	S	73	0.2	S
397	hVH_6_1	hVL_2-14						0	0.4	S
398	**hVH_6_1**	**hVL_2-23**	**2.1**	**0.8**	**60**	**S**	**S**	**69**	**0.4**	**S**
399	hVH_6_1	hVL_3-1		0.5	60	S	S	55	1.4	S
400	hVH_6_1	hVL_3-21	0.4	0.8	60	S	S	66	0.5	S

Table 12 Key:

For relative Fab display, relative Fab expression and relative IgG1 expression, the values illustrate the levels as compared to a control. Higher numbers indicate higher levels.

For Fab thermostability, the numbers 60 and 70 indicate VH/VL pairs which are stable for 45 minutes at 60°C or 70°C at the tested conditions. The number 4 indicates temperature instable pairs and bg (background) indicates low expression levels.

For Fab stability in mouse serum, Fab stability in bovine serum and IgG1 stability in bovine serum, S stands for stable, U for unstable, and bg for background, at the tested conditions.

As described in the previous examples, the predominant VH and VL germline genes and the predominant VH/VL germline gene pairs were identified from the human immune repertoire, then the predominant VH and VL germline protein sequences were analysed *in silico* in order to identify and select variable heavy chain and variable light chain germline protein sequences having favorable biophysical properties. As shown in Table 5, and Figures 2-3, generally, the top 20VH, top 8Vλ and top 12 Vκ were selected for synthesis, combination and subsequent functional analysis. The germline gene sequences were synthesized and then combined in order to generate 400 germline protein pairs that are representative of the abundant germline gene pairs expressed in the human immune repertoire. The 400 VH/VL germline protein pairs were tested for the following properties: a) relative display after phage production and phage ELISA in Fab format; b) relative Fab expression yield after Fab production in E. coli, E. coli cell lysis and ELISA detection of produced Fab; c) temperature stability of Fab after Fab production in E. coli, E. coli cell lysis and ELISA detection of non-denatured Fab after incubation at increased temperatures; d) bovine/mouse serum stability of Fab from E. coli lysates by ELISA detection of non-denatured Fab after incubation in bovine/mouse serum; e) relative human IgG1 expression yield after IgG1 production in mammalian cells and ELISA detection of secreted IgG1 from cell culture supernatants; and f) bovine serum stability of human IgG1 by ELISA detection of non-denatured IgG1 after incubation in bovine/mouse serum.

Using the data provided in Table 12, one of skill in the art could readily identify the germline protein pairs having favorable biophysical properties.

Generally, the germline protein pairs having a threshold value in each functional property were selected for incorporation in the collections. For example, in some embodiments, the germline protein pairs comprising all of the following properties were selected for incorporation into a collection: i) a relative display rate in Fab format comprising a value within the top 75% of Fabs sampled; ii) an expression yield in Fab format of at least 0.4 as compared to Fab VH1-69 VLA_Vl1-40 AYA; iii) thermal stability at 60°C or more for at least 45 minutes in Fab format; iv) stability in bovine or mouse serum in Fab format for greater than ten days at 37°C; v) an

expression yield in IgG format of at least 0.4 as compared to MOR03080; and vi) stability in serum in IgG format for fourteen days at 37°C. Table 32 shows in **bold and underline** the germline protein pairs comprising all of these functional properties.

As described above, however, germline protein pairs having one or more of the functional properties may be selected for incorporation into collections. Here, an aggregate ranking of the 400 germline protein pairs tested was created, so that each germline protein pair could be ranked against the other giving weight to each of the functional properties tested. This allowed the inventors to select one or more germline protein pairs having one or more or all of the listed functional properties. In some embodiments, the collections comprise all of the germline protein pairs having the above characteristics. In some embodiments, the collection comprises the germline protein pairs having the highest aggregate score of the 400 pairs tested. In some embodiments, the germline protein pairs having aggregate scores within the top 10%, top 20%, or top 30% of the 400 pairs tested were selected for incorporation into collections.

Example 9: Further testing of ~100 VH/VL pairs

Of the 400 germline protein pairs tested above (results shown in Table 12), 95 were selected for further testing. The previous testing of the 400 germline protein pairs for display, expression yield, thermal and serum stability acted as a preliminary filter to remove the germline protein pairs that do not have characteristics thought to be favorable for therapeutic development. The goal was to select a sub-group of germline protein pairs having favorable developability characterictics, while at the same time maintaining a high level of diversity within a collection so that the collection can be used to identify developable candidates against any antigen.

Table 12 shows ~60 bold and underlined germline protein pairs which met the thresholds of an embodiment of the disclosure. Of the 95 germline protein pairs selected for further testing, some were chosen because they met the previous criteria, and it was desirable to further test them. Others were chosen, despite not meeting certain thresholds, so that these pairs could be re-evaluated. Again, one of the goals of the present disclosure is to provide a diverse collection that is able to be used to identify antibodies or fragments against any antigen. The 95 germline protein pairs shown in Figures 16-24 were synthesized as described in Example 5. After synthesis and expression in Fab and IgG1 formats, the 95 germline protein pairs were further tested in both Fab and IgG1 formats for the following a) purified Fab expression yield in mg/L (expression culture), b) purified Fab monomeric content (% monomer), c) purified Fab thermal stability in °C, d) purified IgG1 expression yield in mg/L (cell culture), e)

purified IgG1 monomeric content (% monomer), f) purified IgG1 thermal stability in °C, g) IgG1 isoelectric point and h) IgG stress testing with exposure to acid, including differential scanning fluorometry (DSF), absorption, dynamic light scattering and particle staining.

Example 9.1 Purified Fab testing

Fab fragments representing each of the 95 germline protein pairs selected for further testing were expressed in E. coli and purified. Expression of Fab fragments in E. coli TG-1 F- cells was carried out in 500 ml cultures of 2xYT medium supplemented with 0.1% glucose and chloramphenicol. Cultures were shaken until the OD600nm reached 0.5. Fab expression was induced by addition of IPTG (isopropyl-ß-D-thiogalactopyranoside) and further over night cultivation. Cells were harvested and disrupted using lysozyme. His6-tagged (SEQ ID NO: 203) Fab fragments were isolated via IMAC (Bio-Rad, Munich, Germany) and eluted using imidazole. Buffer exchange to 1x Dulbecco´s PBS (pH 7.2, Invitrogen, Darmstadt, Germany) was performed using PD10 columns (GE Healthcare, Munich, Germany). Samples were sterile filtered (0.2 µm).

Example 9.1.1 Purified Fab expression yield determination

The protein concentrations of purified Fab fragments representing each of the 95 germline protein pairs were determined by UV-spectrophotometry (Nanodrop, peqlab, Erlangen, Germany). The extinction coefficient used was 1.538 mL/mg and measured absorbance at 280nm. The results are shown in Figures 16-18.

Example 9.1.2 Purified Fab thermal stability determination

The thermal stability of purified Fab fragments representing each of the 95 germline protein pairs were determined by differential scanning fluorometry (DSF). Differential scanning fluorometry (DSF) is a fluorescence dye based technique that monitors thermal unfolding (melting point) of a protein of interest. Changes in the fluorescence of a hydrophobic dye interacting with the hydrophobic amino acid side-chains of the unfolding protein are monitored over a temperature ramp.

The following materials were used: Sypro Orange fluorescent dye (Sigma, #S5692); iCycler iQ PCR Plates, 96-well (Biorad, #2239441); Microseal B Adhesive Sealer (Biorad #MSB-1001); 96-well Optical Pad (Biorad, #ADR3296); iCycler iQ5 Thermal cycler (Biorad) and Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA).

Diluted Sypro Orange was added to each well of a 96 well iCycler iQ PCR Plate, and the samples were tested at a final concentration of at least 0.1 mg/ml. The iCycler iQ5 Thermal cycler (Biorad) was used for testing. The temperature was scanned from 20 °C to 95 °C at a heating rate of 60 °C/h, and the temperature of unfolding was calculated by analysis of the midpoint of the fluorescence transition. The results are shown in Figures 16-18 in the Purified Fab Thermafluor column.

Example 9.1.3 Purified Fab separation by size exclusion chromatography

The monomer contents (% monomer) of purified Fab fragments representing each of the 95 germline protein pairs were determined by size exclusion chromatography (SEC). SEC was performed on an ÄKTA Purifier System (GE Healthcare Europe GmbH, Freiburg, Germany). For separation a Superdex75 HR 10/30 column was used (GE Healthcare Europe GmbH, Freiburg, Germany). For each sample 10 µl of protein was loaded onto the column, separation was performed at a flow rate of 0.05 ml/min and recorded analyzing the UV absorption at 260 and 280 nm. The running buffer was composed of Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). The results are shown in Figures 16-18.

Example 9.2 IgG1 expression and purification

IgG1s respresenting each of the 95 germline protein pairs selected for further testing were expressed in HKB11 cells. Eukaryotic HKB11 cells were transfected with a 1:1 ratio of IgG heavy and light chain expression vector DNA. Cell culture supernatant was harvested on day 3 to 4 post transfection and subjected to protein A affinity chromatography (MabSelect SURE, GE Healthcare, Munich, Germany). Buffer exchange was performed with 1 x Dulbcecco´s PBS (pH 7.2, Invitrogen, Darmstadt, Germany) and samples were sterile filtered (0.2 µm pore size).

Example 9.2.1 Purified IgG1 expression yield determination

The protein concentrations of purified IgG1s representing each of the 95 germline protein pairs were determined by UV-spectrophotometry (Nanodrop, peqlab, Erlangen, Germany). The extinction coefficient used was 1.369 mL/mg and measured absorbance at 280nm. The results are shown in Figures 16-18.

Example 9.2.2 Purified IgG1 thermal stability determination

IgG1 thermal stability of purified IgG1s was determined by differential scanning fluorometry (DSF) as described in method 9.1.2. The values shown for each IgG represent the

unfolding events that take place within the variable regions of the IgG. The values representing unfolding of the Fc portion are not shown, as they are generally identical for each human IgG1. The results are shown in Figures 16-18.

Example 9.2.3 Purified IgG1 separation by size exclusion chromatography

The monomeric content (% monomer) of purified IgG1 representing each of the 95 germline protein pairs were determined by size exclusion chromatography (SEC). HP-SEC was performed on a Dionex UltiMate 3000 Titanium HPLC system (Dionex Corporation, Germering, Germany) in combination with Wyatt miniDAWN Treos and Wyatt Optilab rEX (Wyatt Technology Europe, Dernbach, Germany). For separation a Tosoh TSK-Gel G3000SWxl column was used (Tosoh Bioscience, Stuttgart, Germany). For each sample 15 µg of protein was loaded onto the column, separation was performed at a flow rate of 0.5 ml/min and recorded analyzing the UV absorption at 280 nm. The running buffer was composed of Gibco D-PBS, pH 7.4 (Invitrogen, Paisley, USA). The results are shown in Figures 16-18.

Example 9.2.4 Purified IgG1 Isoelectric point (pI) calculation

The Isoelectric point of each germline protein pair in IgG1 format was calculated. Methods of determining the pI of a protein are known to one of skill in the art. For example, the following tools can be used: http://www.expasy.org/tools/pi_tool.html; Vector NTI (Invitrogen, Carlsbad, California). The results are shown in Figures 16-18.

Example 9.2.5 Purified IgG1 stress testing with exposure to acid

As a virus inactivation step is standard during the downstream processing (DSP) of Chemistry, Manufacturing and Control (CMC), the ability of the 95 germline protein pairs to withstand acid was tested by lowering the pH and recording aggregation sensitive data for each of the IgG1s. Each of the germline protein pairs was delivered in a 96-deep-well plate format in a concentration of 2 mg/mL. 150 µL of each was transferred into a 96-well plate. Initial characterization was performed by absorption, dynamic light scattering (DLS), differential scanning fluorometry (DSF) measurements and particle staining. The samples were acidified using 1.8 µL 1M Citrate pH 2.3. Samples were neutralized after 2, 5 hours using 1 M Tris pH9.0.

Example 9.2.5 (a) Purified IgG1 Differential Scanning Fluorometry

In order to evaluate the thermal stability before and after exposure to acid of IgG1s respresenting each of the 95 germline protein pairs selected for further testing, differential

scanning fluorometry (DSF) was performed as described in Example 9.1.2. The values shown for each IgG represent the unfolding events that take place within the variable regions of the IgG. The values representing unfolding of the Fc portion are not shown, as they are generally identical for each IgG. If the Tm (apparent melting point) values before and after exposure to acid are equal then the molecular structure of the antibody was either unaffected by the acid or was able to refold efficiently after exposure. The results are shown in Figures 19, 21, and 23.

Example 9.2.5 (b) Purified IgG1 UV/Vis Absorption

In order to identify aggregating samples turbidity was recorded at 320 nm. Turbidity of IgG solutions was assessed before and after acid exposure respresenting each of the 95 germline protein pairs selected for further testing. The results are shown in Figures 19, 21, and 23. Baseline absorbtion was 0.035 extincition units expected for clear solutions. Increase in absorbtion is caused by light scattering which results in increasing absorption. Values above 0.039 are likely to contain aggregates. Values above 0.045 indicate clear presence of aggregates. Values above 0.06 represent critical aggregation levels which were found for molecules with strongly unfavourable stability.

Example 9.2.5 (c) Purified IgG1 Dynamic Light Scattering

In addition, Dynamic Light Scattering (DLS) was performed on each IgG1 respresenting the 95 germline protein pairs selected. Dynamic light scattering (DLS) is a spectroscopic method to assess the hydrodynamic radius of particles in solution. All DLS experiments were performed using a DynaPro Titan cuvette system (Wyatt Technology Europe, Dernbach, Germany).

In case of visible particle contamination after stress testing, the IgGs were centrifuged in order to remove large aggregates. Figures 20, 22 and 24 show the apparent particle radius and polydispersity corresponding to the monomeric IgG1 found in the preparations before and after acid treatment. The data was evaluated according to the calculated radius of the cumulant analysis.In addition to the hydrodynamic radius, the % polydispersity of the preparations was assessed. An increase in polydispersity (> 15 %) indicates potential aggregation of the IgG molecules, leading to heterogeneous particle size distribution. High molecular weight (HMW) particles clearly distinguishable from the IgG (radius > 3-fold) are not listed in the table. All DLS results are shown in Figures 20, 22 and 24.

Example 9.2.5 (d) Purified IgG1 particle staining

In order to evaluate the amount and morphology of visible aggregates, particle staining was performed before and after acid exposure on each IgG1 representing the 95 germline protein pairs selected. The following reagentes were used to filter and stain particles in IgG preparations: Ultrafree-CL 0.22 µm sterile filter (Millipore, #UFC40GV0S); Anti-human lambda light chain, AP conjugated (Sigma #A-2904); Developing agent for AP-conjugates, Fast BCIP/NBT, (Sigma #B-5655); Roti®-ImmunoBlock (Roth #T144.1); Alkaline Phosphatase Stop Solution (Sigma #A5852-100ML); TBS: 0.05 M Tris; 0.15 M NaCl; TBS with 0.1% Tween 20; and 5 M NaCl solution.

The protein solution was filtered through a 0.22 µm filter and the remaining antibody aggregates are subsequently stained using the mouse anti human Fab2 alkaline phosphatase conjugated antibody and a western blot developing agent. The assay was performed according to the manufacturer's manual.The samples were subsequently categorized by visual inspection in range from 1-4, with category 1 representing very low particle content and category 4 representing high particle load of the preparation. All particle staining results are shown in Figures 20, 22 and 24.

Example 9.2.6 Purified IgG1 stress testing with agitation

The ability of antibodies or antibody fragments to resist sheer forces is a helpful criteria as filtration steps cannot be avoided during processing. Therefore, the 95 germline protein pairs were tested in IgG1 format using a glass pearl that was accelerated in a 96 well plate on an orbital shaker at 550 rpm in a deep well plate. 350 µl of each IgG was subjected to this treatment. 150 µL of each was transferred into a 96-well plate. Initial characterization was performed by absorption, dynamic light scattering (DLS), differential scanning fluorometry (DSF) measurements and particle staining.

Example 9.2.6 (a) Purified IgG1 UV/Vis Absorption

In order to identify aggregating samples turbidity was recorded at 320 nm. Turbidity of IgG solutions respresenting each of the 95 germline protein pairs selected for further testing was assessed before and after stress exposure. The results are shown in Figures 49, 51 and 53. Baseline absorbtion was 0.035 extincition units expected for clear solutions. Increase in absorbtion is caused by light scattering which results in increasing absorption. Values above 0.039 are likely to contain aggregates. Values above 0.045 indicate clear presence of

aggregates. Values above 0.06 were found for critical aggregation levels which were found for molecules with strongly unfavourable stability.

Example 9.2.6 (b) Purified IgG1 Differential Scanning Fluorometry

In order to evaluate the thermal stability before and after exposure to acid of IgG1s respresenting each of the 95 germline protein pairs selected for further testing, differential scanning fluorometry (DSF) was performed as described in Example 9.1.2. The values shown for each IgG represent the unfolding events that take place within the variable regions of the IgG. The values representing unfolding of the Fc portion are not shown, as they are generally identical for each human IgG1. The results are shown in Figures 50, 52 and 54.

Example 9.2.6 (c) Purified IgG1 Dynamic Light Scattering

In addition, Dynamic Light Scattering (DLS) was performed on each IgG1 representing the 95 germline protein pairs selected. Dynamic light scattering (DLS) is a spectroscopic method to assess the hydrodynamic radius of particles in solution. All DLS experiments were performed using a DynaPro Titan cuvette system (Wyatt Technology Europe, Dernbach, Germany).

In case of visible particle contamination after stress testing, the IgGs were centrifuged in order to remove large aggregates. Figures 50, 52 and 54 show the apparent particle radius and polydispersity corresponding to the monomeric IgG1 found in the preparations after stress treatment. The data was evaluated according to the calculated radius of the cumulant analysis.In addition to the hydrodynamic radius, the % polydispersity of the preparations was assessed. An increase in polydispersity (> 15 %) indicates potential aggregation of the IgG molecules, leading to heterogeneous particle size distribution. High molecular weight (HMW) particles clearly distinguishable from the IgG (radius > 3-fold) are not listed in the table. All DLS results are shown in Figures 50, 52 and 54.

Example 9.2.6 (d) Purified IgG1 particle staining

In order to evaluate the amount and morphology of visible aggregates, particle staining was performed before and after stress exposure on each IgG1 representing the 95 germline protein pairs selected. The following reagentes were used to filter and stain particles in IgG preparations: Ultrafree-CL 0.22 µm sterile filter (Millipore, #UFC40GV0S); Anti-human lambda light chain, AP conjugated (Sigma #A-2904); Developing agent for AP-conjugates, Fast BCIP/NBT, (Sigma #B-5655); Roti®-ImmunoBlock (Roth #T144.1); Alkaline Phosphatase Stop

Solution (Sigma #A5852-100ML); TBS: 0.05 M Tris; 0.15 M NaCl; TBS with 0.1% Tween 20; and 5 M NaCl solution.

The protein solution was filtered through a 0.22 µm filter and the remaining antibody aggregates are subsequently stained using the mouse anti human Fab2 alkaline phosphatase conjugated antibody and a western blot developing agent. The assay was performed according to the manufacturer's manual.The samples were subsequently categorized by visual inspection in range from 1-4, with category 1 representing very low particle content and category 4 representing high particle load of the preparation. All particle staining results are shown in Figures 50, 52 and 54.

Example 9.2.7 IgG Stress testing cumulative score

In order to help evaluate the stress testing results of both exposure to acid and agitation with glass beads, a scoring system was created so that the germline protein pairs could be compared. Each data point taken in Examples 9.2.5(a-d), results shown in Figures 19-24 and Examples 9.2.6(a-d), results shown in Figures 49-54 was given a score ranging from 0-100 (0, 25, 75 or 100) and the scores were added together to generate a cumulative score. The thermal stability values identified in Examples 9.2.5(a) and 9.2.6 (b) were not given scores.

Figures 55 and 56 show the stress testing scores for the germline protein pairs 1-32 from Examples 9.2.5-9.2.6. Each score is a representation of the raw data points shown in Figures 19, 20, 49 and 50. Figures 19-20 show the response to acid exposure and Figures 49-50 show the response to agitation with glass beads. Figure 56 shows the cumulative score, which is the addition of each of the scores shown in Figures 55 and 56.

Figures 57 and 58 show the stress testing scores for the germline protein pairs 33-64 from Examples 9.2.5-9.2.6. Each score is a representation of the raw data points shown in Figures 21, 22, 51 and 52. Figures 21-22 show the response to acid exposure and Figures 51-52 show the response to agitation with glass beads. Figure 58 shows the cumulative score, which is the addition of each of the scores shown in Figures 57 and 58.

Figures 59 and 60 show the stress testing scores for the germline protein pairs 65-95 from Examples 9.2.5-9.2.6. Each score is a representation of the raw data points shown in Figures 23, 24, 53 and 54. Figures 23-24 show the response to acid exposure and Figures 53-54 show the response to agitation with glass beads. Figure 60 shows the cumulative score, which is the addition of each of the scores shown in Figures 59 and 60.

Example 10: Selection of collection composition

In summary, 400 germline protein pairs were selected, as described in Example 4. These 400 are a representation of the diversity of germline protein pairs that exist in the human immune repertoire. The 400 germline protein pairs were tested as described in Examples 6-7. Of the 400, 95 were further tested as described in Example 9.

The 95 germline protein pairs were compared taking the following factors into consideration: a) Fab display rate; b) Fab expression yield, c) Fab thermal stability; d) Fab serum stability; e) Fab SEC monomeric content (% monomer); f) IgG1 expression yield; g) IgG1 thermal stability; h) IgG1 serum stability; i) IgG1 SEC monomeric content (% monomer); and j) IgG1 isoelectric point (pI). The data for each of these factors are shown in Figures 16-18. These factors correlate well to the developability of therapeutic antibodies.

Fab display rate is an important factor in the selection of antibodies or fragments against an antigen. Fabs displaying at a high rate have a higher likelihood to be exposed to the antigen upon selection. A high display rate of each of the various Fabs makes sure that the full diversity of the collection is exposed to an antigen upon selection. The Fab display rate was identified in Example 6.2, where the reference was an internal standard (HuCAL GOLD reference phage preparation (VH3 kappa + lambda)). The HuCAL GOLD VH3 prep is a high displaying preparation. Fab display rate is an important factor and was useful in narrowing the 400 pairs down to 95 for further testing, but in some embodiments was not considered a determinative factor in the selection of germline protein pairs for incorporation into collections.

Expression yield of both Fab and IgG1 are important as antibodies or fragments selected against an antigen, first must be tested, often *in vitro* or *in vivo* to determine functional activity, then in tox species and finally in humans for clinical trials. It is very important that the antibodies or fragments selected against an antigen can be efficienty expressed in high enough quantity to support all of the various testing required for therapeutic development and for supply of clinical trial and market. The expression yield (mg purified Fab/L of expression culture) of purified Fabs was identified in Example 9.1.1 (results shown in Figures 16-18) and, in an embodiment of the disclosure, a threshold of at least 2.5 mg/L was selected. In other embodiments, other thresholds were selected. The expression yield (mg purified IgG1 /L of cell culture) of purified IgG1 was identified in Example 9.2.1 (results shown in Figures 16-18) and, in an embodiment of the disclosure, a threshold of at least 30.0 mg/L was selected. In other embodiments, other thresholds were selected.

Thermal stability is an important factor as proteins, such as, antibodies, are susceptible to high temperatures, therefore, antibodies capable of withstanding the requirements associated

with the storage and transportation required in order to distribute therapeutics worldwide and have a long shelf life are essential. The thermal stability of purified Fab was determined in Example 9.1.2 (results shown in Figures 16-18) and, in an embodiment of the disclosure, a threshold of at least 70°C was selected. In other embodiments, other thresholds were selected. The thermal stability of purified IgG1 was determined in Example 9.2.2 (results shown in Figures 16-18), the listed value represents the de-stabilization of the variable domains and, in an embodiment, a threshold of at least 73°C was selected. In other embodiments, other thresholds were selected.

Serum stability is an important factor for therapeutic antibodies as therapeutic proteins must maintain efficacy and functional conformation despite being exposed to the serum proteases present in human serum. The serum stability of the germline protein pairs were determined by the methods described in Examples 6.5, and 7.2. Serum stability is important, but was not considered a determinative factor in the selection of germline protein pairs as the assay tended to produce false-negative results in few cases.

Monomeric content (% monomer) as determined by size exclusion chromotagraphy (SEC) is an important factor as it correlates well to aggregation propensity. Aggregation is a common problem in therapeutic protein development, which leads to the inactivation, inhomogeneity and production loss of the protein therapeutic. The monomeric content (% monomer) as determined by size exclusion chromotagraphy (SEC) in both purified Fab and purified IgG1 formats was determined by the methods described in Examples 9.1.3 and 9.2.3 (results shown in Figures 16-18). The monomeric content (% monomer) of purified Fab was determined in Example 9.1.3 and, in an embodiment, a threshold of at least 98% was selected. In other embodiments, other thresholds were selected. The monomeric content (% monomer) of purified IgG1 was determined in Example 9.2.3 and, in an embodiment, a threshold of at least 99% was selected. In other embodiments, other thresholds were selected.

Isoelectric point (pI) is predictive of solubility at a certain pH. When the pH of the solution is significantly different from the pI of a given protein, the protein is soluble. Isoelectric point is important, but in some embodiments was not considered a determinative factor in the selection of germline protein pairs.

In an embodiment of the present disclosure, the thresholds for each criteria were selected as follows: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric

content of purified Fab (as described in Example 9.1.3) of at least 98%; and f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99%. The following germline protein pairs (54) were identified as having these superior functional activities related to developability as each of the following pairs had values equal to or better than these thresholds (data shown in Figures 16-24): VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15

(SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252). Therefore, collections comprising any number of these germline protein pairs can be used to identify developable antibodies or fragments thereof against any antigen.

Additionally, a subset of germline protein pairs were selected based upon a comparison of the stress testing data identified using the methods described in Examples 9.2.5 (a-d), data shown in Figures 19-24, Example 9.2.6 (a-d), data shown in Figures 49-54 and Example 9.2.7, scoring shown in Figures 55-60. The stress testing methods evaluated the 95 germline protein pairs in IgG1 format in order to determine their ability to withstand exposure to acid and agitation with glass beads. 36 germline protein pairs, of an embodiment, were selected as they have additional superior functional properties relevant to developability as they showed strong resistance to acid and agitation stress. An antibody's ability to withstand exposure to acid is an increasingly important factor, as a virus inactivation step is standard during the downstream processing (DSP) of Chemistry, Manufacturing and Control (CMC). The acid treatment step denatures virus capsid proteins, which a virus would use for infection. However, lowering the pH has a destabilizing effect on every protein. Unstable antibodies denature and loose native structure during this step. In the virus activation step, after a defined time, the acid treatment is relieved by neutralization and while the virus capsid proteins stay in an inactive conformation, the processed antibody ideally retains its native structure. The ability of antibodies or antibody fragments to resist sheer forces is a helpful criteria as filtration steps cannot be avoided during processing. These 36 germline protein pairs selected in an embodiment, fulfilled all of the previous threshold functional activities and in addition scored at or above 1225 in the stress testing cumulative score. In an embodiment, the thresholds for each criteria were selected as follows: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric content of purified Fab (as described in Example 9.1.3) of at least 98%; f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99% and g) stress testing cumulative score (as described in Example 9.2.7) of at least 1225. Therefore, embodiments of the present disclosure comprise collections comprising a subset of the fully functional germline protein pairs (36 of the 54) and have additional superior functional properties relevant to developability. In this embodiment, a collection comprises VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ

ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

In another embodiment, the thresholds for each criteria were selected as follows: a) purified Fab expression yield (as described in Example 9.1.1) of at least 2.5 mg/L; b) purified IgG1 expression yield (as described in Example 9.2.1) of at least 30.0 mg/L; c) thermal stability of purified Fab (as described in Example 9.1.2) of at least 70°C; d) thermal stability of purified IgG1 (as described in Example 9.2.2) of at least 73°C; e) monomeric content of purified Fab (as described in Example 9.1.3) of at least 99%; f) monomeric content of purified IgG1 (as described in Example 9.2.3) of at least 99%; g) isoelectric point of purified IgG1 (as described in Example 9.2.4) of at least 8.3; and h) stress testing cumulative score (as described in Example 9.2.7) of at least 1225. In this embodiment, a collection comprises (33 pairs): VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO:

252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252)

In a further embodiment, pairs were added to a collection even though the pairs themselves did not meet all of the thresholds within each criteria, but were added to the collections in order to enhance diversity. In an embodiment, a collection further comprises: VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); and VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256). In this embodiment, a collection comprises (36 pairs): VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO:

213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

Example 11: Beta testing of collections

In order to confirm the effectiveness of the collection design, sub-collections, each comprising one germline protein pair or pools of sub-collections were generated and selected against antigens. The antibodies selected were then tested in both Fab and IgG1 formats for developability characteristics, such as, thermal stability in Fab format, pI in IgG1 format, expression yields in both Fab and IgG1 formats, thermal stability in IgG1 format, and % monomer in IgG1 format as determined by SEC. In addition, in some cases the affinity for the antigen in Fab format was determined.

Collection generation

Sub-collections containing germline protein pairs were synthesized as follows: the FR1-CDR1-FR2-CDR2-FR3 regions from the respective germline protein sequences shown in Figures 25-33 were synthesized by GeneArt (Regensburg, Germany). The VHs were cloned via NheI and SalI and VLs via NdeI and Acc65I into the pJPd1 display vector. CDR-H3 cassettes including the constant FR4 region were inserted via BssHII and XhoI with theoretical diversities ranging between 5.5×10^5 and 1.9×10^{19}. CDR-H3 cassettes with CDR-H3 lengths from 6-17 amino acids were synthesized by Sloning (Martinsried, Germany). CDR-L3 diversity was achieved by introducing either kappy or lambda TRIM cassettes synthesized by ELLA Biotech (Martinsried, Germany) with theoretical diversity ranging between 4.6×10^6 and 2.5×10^9.

Typically 0.25 to 2 µg pJPd1 phagemid DNA of the sub-collections were transformed in E.coli MC1061 F' electrocompetent cells and transformants were collected in TB medium and shaken for at 37°C for 1h. Dilutions of the outgrowth medium were plated on LB/Cam/Gluc. Amplification of the libraries was performed by shaking o/n in appropriate amounts of LB/cam/1%Glu. Library sizes for sub-collections ranged between 4.6×10^8 and 4.4×10^9. The total library size of all sub-collections together is about 1.3×10^{11} members. To analyze the quality of the engineered sub-collections at least 30 clones for each sub-collection were picked

and CDR-L3 and -H3 regions were sequenced to determine correctness and uniqueness of the sequences. The libraries were stored as E. coli glycerol cultures.

Phage displaying the sub-collections in Fab format were prepared as follows. For each library phage preparation 80 ml 2x YT/Cam/Glc medium were inoculated with bacteria from the corresponding library glycerol stock resulting in an OD_{600nm} of 0.2 - 0.3. Cultures were shaken until an OD_{600nm} of 0.45 - 0.55 was reached. Then helper phage was added at a multiplicity of infection of 10 to the bacterial culture followed by an incubation for 45 min at 37°C without shaking and then for 45 min at 37°C shaking at 120 rpm. Bacteria were spun down and helper phage containing supernatant was discarded. Phage-infected bacteria were resuspended in 400 ml 2x YT/CAM/KAN/IPTG medium and incubated overnight at 22°C with shaking at 120 rpm. The next day bacteria from the overnight culture were pelleted and the supernatant containing the Fab-presenting phage was collected. Phage precipitation was performed by adding PEG/NaCl to the phage-containing supernatant. The sample was incubated for at least 30 min on ice. Precipitated phage were spun down and resuspended in PBS. The sample was rotated slowly to obtain a homogeneous suspension and residual bacterial debris was pelleted and discarded. From the phage-containing supernatant the phage were precipitated again using PEG/NaCl. Finally, the phage pellet was resuspended in PBS, transferred to a sterile tube and shaken slowly to obtain a homogeneous suspension. Phage titers were determined by spot titration, ELISA and UV absorbance (Nanodrop) at OD268nm.

Phage titers and display levels of Fab fragments expressed by the tricistronic display vector pJPd1 (shown in Figure 9) and presented on the phage by CysDisplay® (as described in WO01/05950, US 6,753,136, which is incorporated by reference in its entirety) were evaluated for each individual phage preparation by ELISA

Two different antibodies are used for capturing:

(1) The anti-M13 antibody (Amersham #27-9420-01) was used, as it captures phage particles via the major coat protein g8p; therefore, phage titer can be determined.

(2) An anti-Fd antibody (The Binding Site #PC075) was used, which binds to the displayed Fab; therefore, only phage displaying Fabs are captured.

For (1) and (2) separate reference curves are used. A monoclonal anti-M13 (directed against major coat protein of M13 phage, g8p) conjugated to HRP is used as a detection antibody.

The respective capture antibodies were immobilized on 96-well MaxisorpTM plates by dispensing antibody solution for the anti-M13 antibody and for the anti-Fd antibody into different wells, sealing the plate with laminated foil and incubating overnight. The next day, the plates were washed with TBST, and each well was blocked with CTBST.

The starting dilutions of phage supernatants and reference samples (CS) were prepared in CTBST in microtiter plates. The starting dilutions of the phage supernatants for the anti-M13 and anti-Fd antibodies were prepared. The starting dilutions of the reference samples, VH3-23 HuCAL Gold ® l+k VCSM13 and HuCAL PLATINUM pooled Hyperphages kappa and lambda were prepared. Serial dilutions of the phage supernatants were prepared by pre-filling microtiter plates with CTBST and adding phage and pre-filling a second microtiter plate with CTBST, and adding phage. For the reference sample, the starting dilution described above was plated and serial dilutions with both the anti-M13 and anti-Fd antibodies were plated.

Both the phage supernatants and reference samples were transferred for detection as follows. The blocked ELISA plates were washed with TBST. The phage supernatants were transferred from the dilution plates to the coated ELISA plates, incubated at room temperature, and washed with TBST. Anti-M13 peroxidase conjugate (Amersham) diluted in CTBST was added, and incubated for 1- 2 h at room temperature. The Quanta Blu (Pierce) working solution was prepared by mixing 1 part (e.g. 0.5 ml) peroxide solution with 9 parts (e.g. 4.5 ml) substrate solution. The ELISA plates were washed with TBST, the QuantaBlu working solution was added. The fluorescence was measured after an incubation time of ~ 2 min (excitation: 320 nm, emission: 430 nm) and subsequently at intervals of 5 min. The evaluation of the ELISA data was completed as follows: calibration curves were created and the titers of the phage supernatants and control were calculated. For each sample, the titer on anti-Fd was divided by the titer on anti-M13 (anti-pVIII), the resulting ratio was the relative display rate. The results are shown in Table 13.

Table 13

	Framework		Titer (Spot-Titration)		Titer (ELISA)		relative display	
Sub-Library	VH	VL	phageprep I	phageprep II	phageprep I	phageprep II	phageprep I	phageprep II
I8	VH3-23	VK1-39	5.7E+12	2.9E+12	2.9E+13	8.0E+12	**6.1**	**9.3**
I19	VH3-23	VL3-1	6.6E+12	2.2E+12	2.8E+13	9.2E+12	**6.6**	**8.8**

Phage Display Selection against human DKK3, rhErbB4/Her4 Fc fusion, rhFZD-4 Fc fusion and eGFP

Parallel panning strategies with individual sub-collections or pools of sub-collections were performed in order to maximize the chance of identifying diverse binding antibodies with the desired biophysical characteristics. human Dickkopf- 3 (DKK3) (Gene ID 27122), Recombinant human (rh)ErbB4/Her4 (Gene ID 2066)_Fc fusion protein, rhFZD-4 (Gene ID 8322) Fc fusion and eGFP (enhanced green fluorescent protein; sequence provided above) were chosen as model antigens for collection validation. Collection screening was performed in

a M-450 epoxy bead-based solution panning with the respective antigens covalently coupled to magnetic Dynabeads(R) (Dynal/Invitrogen Prod. no. 140.11), described below.

Bead-based solution panning against DKK3

DKK3 and control BSA coated carboxyl-beads (Dynal) were blocked with MPBST at room-temperature (RT) before incubation with pre-adsorbed phages. After several washing steps, bound phage were eluted and amplified by infecting TG1F+ cells for the next round of selection. After 3 rounds of selection, pJPd1 (shown in Figure 9) phagemid DNA was isolated and Fab encoding fragments (modified ompA-VL and modified phoA-Fd) were excised by restriction digestion with XbaI and EcoRI and ligated into the expression vector pJPx1 (shown in Figure 10) and transformed into E. coli TG1 F-. The infected cultures were then plated on large LB/Cam/Gluc plates and allowed to grow over night. Single clones were isolated and tested for Fab expression yield and antigen binding by ELISA. Fab expression was detected by incubating Fab containing cell extracts on a sheep anti-human Fd (The Binding Site Cat. PC075) coated ELISA plate followed by detection with goat anti-human IgG F(ab')2 fragment specific antibody conjugated with Alkaline Phosphatase (AP) (Jackson Cat. 109-055-097). Antigen specificity was tested by screening Fab containing cell extracts on DKK3 coupled-Carboxylbeads and BSA coupled-Carboxylbeads (Dynal) with a fluorometric microvolume assay technology (FMAT®) for bead based assays (Applied Biosystems 8200 Cellular Detection System / PE Biosystems). Primary Hits were defined as Fabs that result in an FMAT mean fluorescence signal of at least 5-fold above the background which was set to a value of 200. Specificity to DKK3 was confirmed in a secondary ELISA with DKK3 as cognate antigen and CD38_Fc as negative control antigen. Heavy and light chain CDR3 region of 63, 43 and 44 clones for the VH3-23/VK1-39, VH3-23/VL3-1 and HuCAL Platinum ® VH3-23/kappa sub-libraries were picked for sequencing in order to estimate the sequence diversity of DKK3 binding antibodies. The sequences of the CDR-H3s and CDR-L3s of selected binders are shown in Figures 86. In total, 31 out of 56 successful sequences (55%), 20 out of 35 sequences (47%) and 17 out of 44 sequences (39%) for the VH3-23/VK1-39, VH3-23/VL3-1 and HuCAL-Pt VH3-23/kappa sublibraries, respectively were different, showing that the constructed libraries contained a diverse repertoire of DKK3 binders. Results are shown in Table 14.

Table 14

Dkk-3					
library	screened	Hits	Hit-rate [%]	picked for Seq.	unique / sequences
I8	**732**	**525**	**72**	**63**	**31 / 56**
I19	**715**	**536**	**75**	**43**	**20 / 35**
HuCAL-Pt VH3-23/k	**736**	**667**	**91**	**44**	**17 / 44**

I8 represents VH3-23/VK1-39, and I19 represents VH3-23/VL3-1.

Bead-based solution panning against rhErbB4/Her4_Fc fusion, rhFZD-4 Fc fusion and eGFP

rhErbB4/Her4_Fc fusion, rhFZD-4 Fc fusion or eGFP and control BSA epoxy M450-beads (Dynal) were blocked with Chemiblocker for 2h at room-temperature (RT) before incubation with pre-adsorbed phages for 2h at RT. After several washing steps, bound phage were eluted and amplified by infecting TG1F+ cells for the next round of selection. After 3 rounds of selection, pJPd1 (shown in Figure 9) phagemid DNA was isolated and Fab encoding fragments (modified ompA-VL and modified phoA-Fd) were amplified by PCR, purified, and digested with XbaI and EcoRI and ligated into the expression vector pJPx1 (shown in Figure 10) and transformed into E. coli TG1 F-. The infected cultures were then plated on large LB/Cam/Gluc plates and allowed to grow overnight. Single clones were isolated and tested for Fab expression yield and antigen binding by ELISA. Fab expression was detected by incubating Fab containing cell extracts on a sheep anti-human Fd (The Binding Site Cat. PC075) coated ELISA plate followed by detection with goat anti-human IgG F(ab')2 fragment specific antibody conjugated with Alkaline Phosphatase (AP) (Jackson Cat. 109-055-097). Antigen specificity was tested by ELISA screening with Fab containing cell extracts on rhErbB4/Her4_Fc antigen, rhFZD-4_Fc antigen or eGFP directly coated on MaxiSorp plates. Primary Hits were defined as Fabs that result in an ELISA signal of at least 5-fold above the background. The results are shown in Figures 61A-D.

Fc-capture panning against ErbB4/Her4_Fc

Three rounds of solid phase Fc-capture panning were performed using human ErbB4/Her4 recombinant Fc-tagged protein immobilized by capturing with goat anti human-IgG Fc specific (Jackson; Cat. 109-005-098) or mouse anti human-IgG Fc specifc (Jackson; Cat. 209-005-098) on Maxisorp plates (Nunc). Prior to each selection round, phages were blocked with 0.1mg/ml human, goat and mouse immunoglobulin in MPBST/BSA. After several washing steps, bound phage were eluted and amplified by infecting TG1F+ cells for the next round of selection. After the third selection round, pJPd1 (shown in Figure 9) phagemid DNA was

isolated and Fab encoding fragments (modified ompA-VL and modified phoA-Fd) were exised by restriction digestion with XbaI and EcoRI and ligated into the expression vector pJPx1(shown in Figure 10) and transformed into TG1F-. The infected cultures were then plated on large LB/Cam/Gluc plates and allowed to grow overnight. Single clones were isolated and tested for Fab expression yield and antigen binding by ELISA. Fab expression was detected by incubating Fab containing cell extracts on a sheep anti-human Fd (The Binding Site Cat. PC075) coated ELISA plate followed by detection with goat anti-human IgG F(ab')2 fragment specific antibody conjugated with Alkaline Phosphatase (AP) (Jackson Cat. 109 055 097). Antigen specificity was tested by ELISA screening with Fab containing cell extracts on ErbB4/Her4_Fc antigen captured via goat anti-human IgG antibody (Jackson; Cat. 109-005-098) coated on MaxiSorp plates. Primary Hits were defined as Fabs that result in an ELISA signal of at least 5-fold above the background. Specificity to ErbB4/Her4_Fc was confirmed in a secondary Fc-capture ELISA with ErbB4/Her4_Fc as cognate antigen and CD38_Fc as negative control antigen.

Heavy and light chain CDR3 regions of 112, 61 and 95 clones for the VH3-23/VK1-39, VH3-23/VL3-1 and HuCAL-Pt VH3-23/kappa sub-libraries, respectively, were sequenced in order to estimate the sequence diversity of ErbB4/Her4_Fc binding antibodies. In total, 31 out of 106 successful sequences (29%), 30 out of 61 sequences (49%) and 14 out of 91 sequences (15%) for the VH3-23/VK1-39, VH3-23/VL3-1 and HuCAL-Pt VH3-23/kappa sub-libraries were different, showing that the constructed libraries contained a diverse repertoire of binders. The sequence diversity is shown in Table 15.

Table 15

Her4_Fc											
library	screened	Hits			Hits*	Hit-rate [%]	picked for CP**				unique ***
		10x Bg	5x Bg	2x Bg			10x Bg	5x Bg	2x Bg	total	
I8	**794**	**112**	**150**	92	**262**	**33**	**86**	**19**	**7**	**112**	**31 / 106**
I19	**1145**	**39**	**7**	15	**46**	**4**	**39**	**7**	**15**	**61**	**30 / 61**
HuCAL-Pt VH3-23/k	**1364**	**922**	**105**	118	**1027**	**75**	**95**	**0**	**0**	**95**	**14 / 91**

* Hits are defined as being reactive at least 5x above background (Bg)

** For compression plates (CP) a few clones were picked that were reactive only 2x above background (Bg)

*** unique sequences per analyzable sequences

I8 represents VH3-23/VK1-39, and I19 represents VH3-23/VL3-1.

Biacore K_D (affinity) determination via antigen capture setup in Fab format

Binding of monomeric Fab fractions (analyzed by analytical SEC; Superdex75, Amersham Pharmacia) to captured antigen was analyzed as follows: On a CM5 chip (Biacore/ GE Healthcare) an appropriate anti-antigen tag capture antibody was covalently immobilized using EDC/NHS chemistry. Kinetic measurements were done by capturing the antigen and subsequent injection of six different Fab concentrations (2^n serial dilution). After each cycle the sensor chip was regenerated. A blank injection of running buffer was used for double referencing. All sensorgrams were fitted using BIA evaluation software 3.2 (Biacore/ GE Healthcare), to determine k_{on} and k_{off} rate constants, which were used to calculate K_D.

The Biacore K_D determinations were performed as follows: Running buffer was PBST (phosphate buffered saline pH 7.2 GIBCO + 0.05% Tween-20). Approx. 400 RU antigen with Fc fusion tag (lot# FYY0310041) were captured using an anti-human Fc antibody (Biacore/ GE Healthcare). Fab concentrations ranging from 15.6 to 500 nM were used with a flow rate of 20 µl/min, an injection time of 30 s and a dissociation time of 100 s. Regeneration of the surface was done with 2 injections à 15 µl 3 M $MgCl_2$ reagent. The results are shown in Figure 38.

Developability Testing of Antibodies and antibody fragments identified against DKK3, rhErbB4/Her4 Fc fusion, rhFZD-4 Fc fusion and eGFP

The antibodies or fragments specific for the antigens were tested in both Fab and IgG1 formats for developability characteristics, such as, thermal stability in Fab format, affinity in Fab format, pI in IgG1 format, expression yield in both Fab and IgG formats, thermal stability in IgG1 format, and % monomer in IgG1 format as determined by SEC. The serum stability in IgG1 format was tested as described in Example 7.2. The thermal stability testing in Fab and IgG1 formats was completed as described in Examples 9.1.2 and 9.2.2. The pI in IgG1 format was completed as described in Example 9.2.4. The expression yield in IgG1 format was completed as described in Example 9.2.1. The % monomer in IgG1 format as determined by SEC was completed as described in Example 9.2.3. The results are shown in Figures 37-39, 45-48 and 62.

Again, the inventors believe that there is a high correlation between the input (antibody collection used for selection against an antigen) and output (antibodies identified as specific for the antigen) regarding the tested functional properties. Therefore, the collections of the invention comprise antibodies or fragments that comprise, in part, the same amino acid sequences as the constructs tested, for example, the framework regions and/or complementarity determining regions. The CDR3s are diversified. Since, in an aspect, the

collections comprise the amino acid sequences, or the nucleic acids encoding them, of the tested constructs it is believed that the collections comprise antibodies or fragments having the same superior functional properties related to developabiltiy as the constructs tested in Example 9. Therefore, it is expected that many of the antibodies or fragments subsequently selected against an antigen will also have the same superior functional properties relevant to developability.

The data shown in Figures 37-39, 45-48 and 62A-C support this conclusion. Figure 39 shows the Fabs selected against DKK3 or ErbB4/Her4_Fc antigen from collections of the invention and how the Fabs have a similar thermal stability as the control, which was the construct originally tested as described in Example 9. In addition, Figures 45-48 show the IgGs specific for DKK3 or ErbB4/Her4_Fc antigen that were selected from the collections of the invention and how the IgGs have similar isoelectric points (pI), thermal stability, expression yield and monomeric content as the controls, which were the constructs originally tested as described in Example 9. Figures 62A-C shows IgGs selected against rhErbB4/Her4_Fc fusion, rhFZD-4 Fc fusion and eGFP and how the IgGs have similar isoelectric points (pI), expression yield, thermal stability, and monomeric content as the controls, which were the constructs originally tested as described in Example 9.

Overall, this shows that the collections of the invention contain antibodies or fragments having superior properties relevant to developability and supports the inventors' hypothesis that the input, collections using sequences, for example, framework regions and/or complementarity determining regions from the germline protein pairs tested and shown to have superior functional properties, correlates well to the output, antibodies or fragments selected against any antigen having the same superior functional properties related to development.

It is to be understood that the description, specific examples and data, while indicating exemplary embodiments, are given by way of illustration and are not intended to limit the present invention. Various changes and modifications within the present invention will become apparent to the skilled artisan from the discussion, disclosure and data contained herein, and thus are considered part of the invention.

We Claim

1. A collection of synthetic antibodies or functional fragments thereof, wherein at least 50% of the antibodies or functional fragments comprise variable heavy chain and variable light chain pairs, wherein the framework regions of said variable heavy chain and variable light chain pairs consist of twenty five or more germline protein sequences selected from the variable heavy chain and variable light chain pairs having the sequences as defined in figures 25-36: VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO:

250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

2. A collection of synthetic antibodies or functional fragments thereof according to claim 1, wherein the framework regions of the variable heavy chain and variable light chain pairs consist of thirty or more germline protein sequences selected from the variable heavy chain and variable light chain pairs having the sequences as defined in figures 25-36: VH1-18 (SEQ ID NO: 204)/VK1-39 (SEQ ID NO: 236); VH1-18 (SEQ ID NO: 204)/VK3-15 (SEQ ID NO: 238); VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-46 (SEQ ID NO: 205)/VL3-21 (SEQ ID NO: 257); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-16 (SEQ ID NO: 234); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK1-39 (SEQ ID NO: 236); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VK1-05 (SEQ ID NO: 230); VH3-11 (SEQ ID NO: 208)/VK1-39 (SEQ ID NO: 236); VH3-11 (SEQ ID NO: 208)/VK3-15 (SEQ ID NO: 238); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-16 (SEQ ID NO: 234); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-40 (SEQ ID NO: 250); VH3-15 (SEQ ID NO: 209)/VL1-47 (SEQ ID NO: 251); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-15 (SEQ ID NO: 209)/VL2-14 (SEQ ID NO: 254); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-21 (SEQ ID NO: 210)/VK1-27 (SEQ ID NO: 235); VH3-21 (SEQ ID NO: 210)/VL2-11 (SEQ ID NO: 253); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-30 (SEQ ID NO: 212)/VK3-20 (SEQ ID NO: 239); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK1-27 (SEQ ID NO: 235); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH3-74 (SEQ ID NO: 214)/VL1-51 (SEQ ID NO: 252); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-15 (SEQ ID NO: 238); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

3. A collection according to any one of the preceding claims, wherein the framework regions of the variable heavy chain and variable light chain pairs consist of germline protein sequences of the variable heavy chain and variable light chain pairs having the sequences as defined in figures 25-36: VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO:

205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3-07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK1-39 (SEQ ID NO: 236); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-23 (SEQ ID NO: 211)/VL2-23 (SEQ ID NO: 255); VH3-23 (SEQ ID NO: 211)/VL3-1 (SEQ ID NO: 256); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

4. A collection of synthetic antibodies or functional fragments thereof according to claim 3, wherein the framework regions of the variable heavy chain and variable light chain pairs consist of the germline protein sequences of the variable heavy chain and variable light chain pairs having the sequences as defined in figures 25-36: VH1-18 (SEQ ID NO: 204)/VK3-20 (SEQ ID NO: 239); VH1-46 (SEQ ID NO: 205)/VK3-15 (SEQ ID NO: 238); VH1-46 (SEQ ID NO: 205)/VL1-51 (SEQ ID NO: 252); VH1-69*01 (SEQ ID NO: 206)/VL1-51 (SEQ ID NO: 252); VH3-07 (SEQ ID NO: 207)/VK1-12 (SEQ ID NO: 233); VH3-07 (SEQ ID NO: 207)/VK1-27 (SEQ ID NO: 235); VH3 07 (SEQ ID NO: 207)/VK3-15 (SEQ ID NO: 238); VH3-07 (SEQ ID NO: 207)/VL1-47 (SEQ ID NO: 251); VH3-07 (SEQ ID NO: 207)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL1-40 (SEQ ID NO: 250); VH3-11 (SEQ ID NO: 208)/VL1-47 (SEQ ID NO: 251); VH3-11 (SEQ ID NO: 208)/VL1-51 (SEQ ID NO: 252); VH3-11 (SEQ ID NO: 208)/VL2-23 (SEQ ID NO: 255); VH3-15 (SEQ ID NO: 209)/VK1-05 (SEQ ID NO: 230); VH3-15 (SEQ ID NO: 209)/VK1-06 (SEQ ID NO: 231); VH3-15 (SEQ ID NO: 209)/VK1-12 (SEQ ID NO: 233); VH3-15 (SEQ ID NO: 209)/VK1-27 (SEQ ID NO: 235); VH3-15 (SEQ ID NO: 209)/VK3-11 (SEQ ID NO: 237); VH3-15 (SEQ ID NO: 209)/VL1-51 (SEQ ID NO: 252); VH3-21 (SEQ ID NO: 210)/VK1-12 (SEQ ID NO: 233); VH3-23 (SEQ ID NO: 211)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VK3-15 (SEQ ID NO: 238); VH3-53 (SEQ ID NO: 213)/VL2-11 (SEQ ID NO: 253); VH3-74 (SEQ ID NO: 214)/VK1-05 (SEQ ID NO: 230); VH3-74 (SEQ ID NO: 214)/VK1-06 (SEQ ID NO: 231); VH3-74 (SEQ ID NO: 214)/VK1-12 (SEQ ID NO: 233); VH3-74 (SEQ ID NO: 214)/VK3-20 (SEQ ID NO: 239); VH5-51 (SEQ ID NO: 215)/VK1-39 (SEQ ID NO: 236); VH5-51 (SEQ ID NO: 215)/VL1-40 (SEQ ID NO: 250); VH5-51 (SEQ ID NO: 215)/VL1-51 (SEQ ID NO: 252); VH6-1 (SEQ ID NO: 216)/VK1-09 (SEQ ID NO: 232); VH6-1 (SEQ ID NO: 216)/VK3-20 (SEQ ID NO: 239) and VH6-1 (SEQ ID NO: 216)/VL1-51 (SEQ ID NO: 252).

5. The collection according to any one of the preceding claims, wherein said antibodies or functional fragments thereof comprise one or more complementarity determining regions consisting of germline protein sequences from the respective variable heavy chain and variable light chain pairs.

6. The collection according to any one of the preceding claims, wherein said antibodies or functional fragments thereof comprise CDR1 regions consisting of germline protein sequences from the respective variable heavy chain and variable light chain pairs.

7. The collection according to any one of the preceding claims, wherein said antibodies or functional fragments thereof comprise CDR2 regions consisting of germline protein sequences from the respective variable heavy chain and variable light chain pairs.

8. The collection according to any one of the preceding claims, wherein said antibodies or functional fragments thereof comprise one or more complementarity determining regions comprising amino acid modifications that remove potential post translational modification sites.

9. The collection according to claim 8, wherein said antibodies or functional fragments thereof comprise one or more heavy chain complementarity determining regions comprising the complementary determining region sequences from the respective variable heavy chains depicted in figures 34-36.

10. The collection according to claim 9, wherein said antibodies or functional fragments thereof comprise HCDR1 regions from the respective HCDR1 region depicted in figure 34.

11. The collection according to claims 9 or 10, wherein said antibodies or functional fragments thereof comprise HCDR2 regions from the respective HCDR2 region depicted in figure 35.

12. The collection according to any one of the preceding claims, wherein said antibodies or functional fragments thereof comprise a FR4 region selected from the group consisting of JH4, Jκ1, and Jλ2/3 the sequences thereof being defined in figure 40.

13. The collection according to any one of the preceding claims, wherein said antibodies or functional fragments thereof comprise a diversified HCDR3 region.

14. The collection according to any one of the preceding claims, wherein said antibodies or functional fragments thereof comprise a diversified LCDR3 region.

15. The collection according to any one of the proceeding claims, wherein the collection comprises at least 1 x 10^4 antibodies or functional fragments thereof.

16. The collection according to any one of the proceeding claims, wherein said antibodies are selected from the groups consisting of human IgG1, IgG2, IgG3, IgG4, IgA, IgE, IgM and IgD.

17. The collection according to any one of the proceeding claims, wherein said functional fragments of said antibodies are selected from the group consisting of Fab, F(ab')2, Fab', Fv, and scFv.

18. A collection of nucleic acids encoding a collection of antibodies or functional fragments thereof according to any one of the preceding claims.

19. A vector comprising the collection of nucleic acids according to claim 18.

20. A recombinant host cell comprising the nucleic acids of claim 18 or the vector of claim 19.

21. A method of identifying an antibody or antibody fragment specific for an antigen, comprising: (a) contacting the antigen with a collection of antibodies or functional fragments thereof according to claim 1, and (b) selecting one or more antibodies or antibody fragments that bind to said antigen.